



# **STIC Search Report**

## **Biotech-Chem Library**

STIC Database Tracking Number 122814

TO: David Lukton  
Location: REM-3B75/3C70  
Art Unit: 1653  
Tuesday, May 25, 2004

Case Serial Number: 09/594978

From: Toby Port  
Location: Biotech-Chem Library  
Remsen 1A59  
Phone: 571-272-2523

[toby.port@uspto.gov](mailto:toby.port@uspto.gov)

### **Search Notes**

Dear Examiner Lukton,

Here are the results of your search.  
Please feel free to contact me if you have any questions.

Toby Port

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

on on: May 24, 2004, 17:31:37 ; Search time 41.4286 Seconds  
(without alignments)

27.280 Million cell updates/sec

itle: US-09-594-978A-2

ffect score: 19

equences: 1 VAEF 4

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

otal number of hits satisfying chosen parameters: 1586107

inimum DB seq length: 0

aximum DB seq length: 2000000000

set-processing: Minimum Match 0%

Listing first 1000 summaries

atabase : A\_Geneseq\_29Jan04.\*

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	19	100.0	6	5	AAW47151 S chrysom
2	19	100.0	7	3	ABG07871 A beta-se
3	19	100.0	8	3	ABG07872 A beta-se
4	19	100.0	8	4	ABB56241 Vascular
5	19	100.0	8	4	AAU28720 DPI trypt
6	19	100.0	8	4	AAU25114 Schizophr
7	19	100.0	8	4	AAU26368 Depressio
8	19	100.0	8	4	AAU15458 Schizophr
9	19	100.0	9	2	AAx78909 MAGE 3 10
10	19	100.0	9	5	ABx77871 A Beta-am
11	19	100.0	10	5	ABG06593 Beta-secr
12	19	100.0	12	5	ABG06592 Beta-secr
13	19	100.0	12	5	ABG78404 Memapsin
14	19	100.0	12	7	ABD35467 Escherich
15	19	100.0	13	3	ABG07889 A beta-se
16	19	100.0	13	4	AAx99276 Vaccine r
17	19	100.0	14	3	AAx07888 A peptide
18	19	100.0	14	3	AAx07888 Human pep
19	19	100.0	15	6	AAE32223 Arabidops
20	19	100.0	20	2	AAW42944 Immunogen
21	19	100.0	20	2	AAW42943 Immunogen
22	19	100.0	21	4	AAE69462 Synthetic
23	19	100.0	21	4	AAE69463 Synthetic
24	19	100.0	25	2	AAW42946 Immunogen
25	19	100.0	25	4	AAE69465 Synthetic

26	19	100.0	28	2	AAW17832	AAW17832 Human mev
27	19	100.0	28	4	AAW33910	AAW33910 Peptide #
28	19	100.0	28	4	ABG55464	ABG55464 Human liv
29	19	100.0	28	5	ABG43601	ABG43601 Human pep
30	19	100.0	30	2	AAW62760	AAW62760 Streptoco
31	19	100.0	32	2	AAW82654	AAW82654 Human imm
32	19	100.0	33	4	AAO09024	AAO09024 Human pol
33	19	100.0	35	2	AAW77953	AAW77953 Antigenic
34	19	100.0	35	2	AAW46146	AAW46146 Predicted
35	19	100.0	35	2	AAW51751	AAW51751 H. influe
36	19	100.0	35	2	AAW55788	AAW55788 Human aqu
37	19	100.0	35	2	AAW54105	AAW54105 Tbp2 anti
38	19	100.0	35	2	AAW94320	AAW94320 Human aqu
39	19	100.0	35	3	AAW80448	AAW80448 H. influe
40	19	100.0	36	2	AAW27782	AAW27782 UDP-N-ace
41	19	100.0	37	4	AAW99833	AAW99833 Human exc
42	19	100.0	37	4	AAW42648	AAW42648 Human kid
43	19	100.0	43	4	AAW17544	AAW17544 Peptide #
44	19	100.0	43	4	ABB31362	ABB31362 Peptide #
45	19	100.0	43	4	ABB21904	ABB21904 Protein #
46	19	100.0	43	4	ABG51414	ABG51414 Human liv
47	19	100.0	43	4	AAW05214	AAW05214 Peptide #
48	19	100.0	47	3	AAW27178	AAW27178 RSV parti
49	19	100.0	48	4	AAW18541	AAW18541 Peptide #
50	19	100.0	48	4	ABB37578	ABB37578 Peptide #
51	19	100.0	48	4	AAW30996	AAW30996 Peptide #
52	19	100.0	48	4	ABB22868	ABB22868 Protein #
53	19	100.0	48	4	AAW70687	AAW70687 Human bon
54	19	100.0	48	4	ABG52389	ABG52389 Human liv
55	19	100.0	48	4	AAW06107	AAW06107 Peptide #
56	19	100.0	53	4	AAO10565	AAO10565 Human pol
57	19	100.0	53	4	AAO05247	AAO05247 Human pol
58	19	100.0	56	4	AAW87390	AAW87390 Human imm
59	19	100.0	56	6	ABP97126	ABP97126 Human mat
60	19	100.0	56	6	ABG76312	ABG76312 Human mat
61	19	100.0	58	2	AAW55888	AAW55888 Rat PCL2.
62	19	100.0	58	5	ABP04854	ABP04854 Human ORF
63	19	100.0	59	3	AAW60085	AAW60085 Arabidops
64	19	100.0	59	3	AAW60683	AAW60683 Arabidops
65	19	100.0	60	4	ABW42044	ABW42044 Peptide #
66	19	100.0	60	4	AAW35846	AAW35846 Peptide #
67	19	100.0	60	4	AAW75737	AAW75737 Human bon
68	19	100.0	60	4	AAW62925	AAW62925 Human bra
69	19	100.0	60	4	ABG57475	ABG57475 Human liv
70	19	100.0	62	2	AAV17262	AAV17262 HLF domai
71	19	100.0	62	4	AAW91247	AAW91247 Human imm
72	19	100.0	62	6	ABW65667	ABW65667 Propionib
73	19	100.0	64	4	AAO09517	AAO09517 Human pol
74	19	100.0	67	2	AAW48352	AAW48352 Human ORF
75	19	100.0	68	5	ABP31047	ABP31047 Human ORF
76	19	100.0	69	3	AAW32809	AAW32809 Zea may
77	19	100.0	69	3	AAW21161	AAW21161 Zea may
78	19	100.0	69	4	ABW42991	ABW42991 Peptide #
79	19	100.0	69	4	AAW36821	AAW36821 Peptide #
80	19	100.0	69	4	AAW78713	AAW78713 Human bon
81	19	100.0	69	4	AAW67671	AAW67671 Human bra
82	19	100.0	69	4	ABG58405	ABG58405 Human liv
83	19	100.0	69	4	AAW91842	AAW91842 C glutami
84	19	100.0	71	5	ABP08064	ABP08064 Human ORF
85	19	100.0	71	5	AAE15877	AAE15877 Hepatitis
86	19	100.0	71	6	AAW35824	AAW35824 Hepatitis
87	19	100.0	73	4	AAU53554	AAU53554 Propionib
88	19	100.0	73	4	AAU66465	AAU66465 Propionib
89	19	100.0	73	5	ABP06455	ABP06455 Human ORF
90	19	100.0	73	6	ABW50073	ABW50073 Propionib
91	19	100.0	73	6	ABW62984	ABW62984 Propionib
92	19	100.0	74	5	AAE15876	AAE15876 Hepatitis
93	19	100.0	75	4	ABG00769	ABG00769 Novel hum
94	19	100.0	76	4	AAW45469	AAW45469 Propionib
95	19	100.0	76	6	ABW41988	ABW41988 Propionib
96	19	100.0	79	5	ABP62931	ABP62931 Human pol
97	19	100.0	83	6	ABP79418	ABP79418 N. gonorr
98	19	100.0	84	5	ABP66623	ABP66623 Human bre

99	19	100.0	84	5	ABP33668	Human ORF
100	19	100.0	86	4	AAU50485	Propionib
101	19	100.0	86	6	ABM47004	Propionib
102	19	100.0	87	4	AAU505209	Human pol
103	19	100.0	87	4	AAU43264	Propionib
104	19	100.0	87	4	AAU66125	Propionib
105	19	100.0	87	6	ABM62644	Propionib
106	19	100.0	87	6	ABM39783	Propionib
107	19	100.0	88	4	AAU56968	Propionib
108	19	100.0	88	6	ABM53487	Propionib
109	19	100.0	88	7	ABM53487	Propionib
110	19	100.0	88	7	ABM53487	Propionib
111	19	100.0	89	3	AAU16761	Bacteriop
112	19	100.0	89	4	AAU35652	Haemophil
113	19	100.0	89	4	AAU35631	Haemophil
114	19	100.0	89	5	ABM54895	Lactococc
115	19	100.0	89	6	ABU40850	Protein e
116	19	100.0	89	6	ABU30565	Protein e
117	19	100.0	89	6	ABU30566	Protein e
118	19	100.0	89	6	ABU30566	Protein e
119	19	100.0	90	3	AAU30336	Protein e
120	19	100.0	90	3	AAU30336	Protein e
121	19	100.0	90	3	AAU30336	Protein e
122	19	100.0	90	3	AAU30336	Protein e
123	19	100.0	90	3	AAU30336	Protein e
124	19	100.0	90	3	AAU30336	Protein e
125	19	100.0	90	3	AAU30336	Protein e
126	19	100.0	90	3	AAU30336	Protein e
127	19	100.0	90	3	AAU30336	Protein e
128	19	100.0	90	3	AAU30336	Protein e
129	19	100.0	90	3	AAU30336	Protein e
130	19	100.0	90	3	AAU30336	Protein e
131	19	100.0	90	3	AAU30336	Protein e
132	19	100.0	90	3	AAU30336	Protein e
133	19	100.0	90	3	AAU30336	Protein e
134	19	100.0	90	3	AAU30336	Protein e
135	19	100.0	90	3	AAU30336	Protein e
136	19	100.0	90	3	AAU30336	Protein e
137	19	100.0	90	3	AAU30336	Protein e
138	19	100.0	90	3	AAU30336	Protein e
139	19	100.0	90	3	AAU30336	Protein e
140	19	100.0	90	3	AAU30336	Protein e
141	19	100.0	90	3	AAU30336	Protein e
142	19	100.0	90	3	AAU30336	Protein e
143	19	100.0	90	3	AAU30336	Protein e
144	19	100.0	90	3	AAU30336	Protein e
145	19	100.0	90	3	AAU30336	Protein e
146	19	100.0	90	3	AAU30336	Protein e
147	19	100.0	90	3	AAU30336	Protein e
148	19	100.0	90	3	AAU30336	Protein e
149	19	100.0	90	3	AAU30336	Protein e
150	19	100.0	90	3	AAU30336	Protein e
151	19	100.0	90	3	AAU30336	Protein e
152	19	100.0	90	3	AAU30336	Protein e
153	19	100.0	90	3	AAU30336	Protein e
154	19	100.0	90	3	AAU30336	Protein e
155	19	100.0	90	3	AAU30336	Protein e
156	19	100.0	90	3	AAU30336	Protein e
157	19	100.0	90	3	AAU30336	Protein e
158	19	100.0	90	3	AAU30336	Protein e
159	19	100.0	90	3	AAU30336	Protein e
160	19	100.0	90	3	AAU30336	Protein e
161	19	100.0	90	3	AAU30336	Protein e
162	19	100.0	90	3	AAU30336	Protein e
163	19	100.0	90	3	AAU30336	Protein e

245	19	100.0	164	7	ADB87765	Ade87765 Human lun	318	19	100.0	199	4	AAB68971	Aab68971 Sheep mit
246	19	100.0	167	3	ADB56403	Aab56403 Human pro	319	19	100.0	199	4	AAB68972	Aab68972 Human mit
247	19	100.0	167	6	ABF79188	Abf79188 N. gonorr	320	19	100.0	199	4	AAB31894	Aab31894 Amino aci
248	19	100.0	168	3	ACG38110	Agc38110 Arabidops	321	19	100.0	199	4	AAB31892	Aab31892 Amino aci
249	19	100.0	170	2	ABG26260	Abg26260 Novel hum	322	19	100.0	199	4	AAB31893	Aab31893 Amino aci
250	19	100.0	172	2	AAW40216	Aaw40216 Mouse Al.	323	19	100.0	199	4	AB83265	Aab83265 A nidulan
251	19	100.0	175	2	AAW05531	Aaw05531 Human ant	324	19	100.0	199	4	AB83265	Aab83265 A nidulan
252	19	100.0	175	2	AAW31531	Aaw31531 Human ant	325	19	100.0	199	4	AB83265	Aab83265 A nidulan
253	19	100.0	175	2	AAW20421	Aaw20421 H. pylori	326	19	100.0	199	5	AAW47317	Aam47317 Human nuc
254	19	100.0	175	5	AAU12115	Aau12115 Human bcl	327	19	100.0	199	7	ADD46946	Add46946 Human pro
255	19	100.0	175	6	ABU89728	Abu89728 Protein d	328	19	100.0	199	7	ADD46942	Add46942 Human pro
256	19	100.0	175	7	ADB25780	Adb25780 Human pro	329	19	100.0	200	4	AA79735	Aab79735 Coryneb
257	19	100.0	176	4	AAW14209	Aam14209 Peptide #	330	19	100.0	200	4	AA79735	Aab79735 Coryneb
258	19	100.0	176	4	ABB33156	Abb33156 Peptide #	331	19	100.0	201	3	AA79735	Aab79735 Coryneb
259	19	100.0	176	4	AAW26619	Aam26619 Peptide #	332	19	100.0	201	3	AA79735	Aab79735 Coryneb
260	19	100.0	176	4	ABB27984	Abb27984 Human pep	333	19	100.0	201	4	AA79735	Aab79735 Coryneb
261	19	100.0	176	4	ABB18621	Abb18621 Protein #	334	19	100.0	201	5	AAE15548	Aae15548 Human sec
262	19	100.0	176	4	AAW66340	Aam66340 Human bon	335	19	100.0	201	7	ADE57400	Ade57400 Human pro
263	19	100.0	176	4	AAW53952	Aam53952 Human bra	336	19	100.0	201	7	ADE56860	Ade56860 Human pro
264	19	100.0	176	4	ABG48006	Abg48006 Human liv	337	19	100.0	201	7	ADE56856	Ade56856 Human pro
265	19	100.0	176	4	AAW01941	Aam01941 Peptide #	338	19	100.0	202	3	AG08264	Ag08264 Arabidops
266	19	100.0	176	5	ABG35988	Abg35988 Human pep	339	19	100.0	202	5	AB47473	Ab47473 Listeria
267	19	100.0	177	6	ABP80373	Abp80373 N. gonorr	340	19	100.0	202	6	ABU32939	Abu32939 Protein e
268	19	100.0	177	6	ABP79196	Abp79196 Human can	341	19	100.0	203	4	AA63333	Aab63333 Human bre
269	19	100.0	178	3	AA343469	Aab43469 Human can	342	19	100.0	203	7	ADC94985	Adc94985 E. faeciu
270	19	100.0	178	3	AA544140	Aab54140 Human pan	343	19	100.0	204	5	ABG61835	Abg61835 Prostata
271	19	100.0	178	3	AA544526	Aab44526 Virulence	344	19	100.0	205	2	AAW13020	Aaw13020 Human MAD
272	19	100.0	178	4	AB62070	Ab62070 Drosophil	345	19	100.0	205	4	AAU27824	Aau27824 Human ful
273	19	100.0	178	5	ABP41446	Abp41446 Human ova	346	19	100.0	205	4	AAU27824	Aau27824 Human ful
274	19	100.0	178	5	ABP54478	Abp54478 Pasteurel	347	19	100.0	205	5	ABP65838	Abp65838 Bifidobac
275	19	100.0	180	3	AGL10317	Ag10317 Arabidops	348	19	100.0	205	6	ABR58559	AbR58559 Human can
276	19	100.0	180	3	AGG61314	Agg61314 Arabidops	349	19	100.0	205	6	ABU56531	Abu56531 Lung canc
277	19	100.0	180	6	AD354552	Ad354552 Acinetoba	350	19	100.0	205	6	ABP75507	Abp75507 Human sec
278	19	100.0	181	3	AGS9111	AgS9111 Arabidops	351	19	100.0	205	6	ABP55379	Abp55379 Human col
279	19	100.0	182	4	AA311895	Aab31895 Amino aci	352	19	100.0	206	7	ABW73792	Abw73792 DNA clone
280	19	100.0	182	4	AA682527	AA682527 S. epider	353	19	100.0	208	6	ABU16876	Abu16876 Protein e
281	19	100.0	184	7	ADC00843	AdC00843 Enterohae	354	19	100.0	209	6	ABU44637	Abu44637 Protein e
282	19	100.0	185	4	AAU42124	Aau42124 Propionib	355	19	100.0	210	4	AAU23550	Aau23550 Novel hum
283	19	100.0	185	6	ABW38643	Abw38643 Propionib	356	19	100.0	212	6	ABU18781	Abj18781 Pseudomon
284	19	100.0	186	4	AAU59747	Aau59747 Propionib	357	19	100.0	213	2	AAU14938	Aay14938 Amino aci
285	19	100.0	186	6	ABW56266	Abw56266 Propionib	358	19	100.0	213	3	AG30715	Ag30715 Arabidops
286	19	100.0	186	6	ABW56266	Abw56266 Propionib	359	19	100.0	213	3	AG30715	Ag30715 Arabidops
287	19	100.0	187	3	AAU72094	Aay72094 Human ser	360	19	100.0	213	3	AG23888	Ag23888 Arabidops
288	19	100.0	187	5	ABG30791	Abg30791 Human gen	361	19	100.0	213	3	AG24431	Ag24431 Arabidops
289	19	100.0	187	5	ABE17922	Aae17922 Human ser	362	19	100.0	213	3	AG24431	Ag24431 Arabidops
290	19	100.0	188	3	AGG05006	AgG05006 Arabidops	363	19	100.0	213	4	AA883228	AA883228 Human FAT
291	19	100.0	188	4	AAW00088	Aaw0088 Alpha-glu	364	19	100.0	213	5	ABG70190	Abg70190 Human pre
292	19	100.0	188	6	ABU50253	Abu50253 Protein e	365	19	100.0	214	4	AG75113	Ag75113 Human col
293	19	100.0	188	6	ABU48662	Abu48662 Protein e	366	19	100.0	214	4	AG75113	Ag75113 Human col
294	19	100.0	189	4	ABW52817	Abw52817 Escherich	367	19	100.0	214	6	AAU48893	Aau48893 Propionib
295	19	100.0	190	2	AAW77092	Aaw77092 Human RNA	368	19	100.0	215	4	AAU48893	Aau48893 Propionib
296	19	100.0	190	2	AAW77092	Aaw77092 Human RNA	369	19	100.0	215	4	AAU48893	Aau48893 Propionib
297	19	100.0	190	4	AAW38905	Aam38905 Human pol	370	19	100.0	216	3	AAW24430	AAW24430 Arabidops
298	19	100.0	191	3	AB41183	Ab41183 Human ORF	371	19	100.0	217	3	AAW24430	AAW24430 Arabidops
299	19	100.0	191	6	AAU49757	Aau49757 Protein e	372	19	100.0	217	6	ABO00751	AbO00751 Polypepti
300	19	100.0	192	3	AGG10316	AgG10316 Arabidops	373	19	100.0	218	3	AG20462	Ag20462 Arabidops
301	19	100.0	192	3	AGG10316	AgG10316 Arabidops	374	19	100.0	218	4	AAU48893	Aau48893 Propionib
302	19	100.0	192	4	AA797936	Aab797936 Coryneb	375	19	100.0	218	6	ABU50613	Abu50613 Protein e
303	19	100.0	192	4	AA797936	Aab797936 Coryneb	376	19	100.0	218	6	ABU50613	Abu50613 Protein e
304	19	100.0	193	4	AA800163	Aab80163 Coryneb	377	19	100.0	219	3	AAW74257	AAW74257 Neisseria
305	19	100.0	193	5	AAU72563	Aau72563 Coryneb	378	19	100.0	219	3	AAW74257	AAW74257 Neisseria
306	19	100.0	194	2	AAU72563	Aau72563 Coryneb	379	19	100.0	219	7	ADD27149	Add27149 Human adi
307	19	100.0	194	6	ABW70654	Abw70654 Phototrab	380	19	100.0	220	7	ADD27149	Add27149 Human adi
308	19	100.0	194	6	ABW70654	Abw70654 Phototrab	381	19	100.0	221	4	AAU40691	Aau40691 Human pol
309	19	100.0	195	3	AGG19885	Agg19885 Arabidops	382	19	100.0	221	4	ABG29763	Abg29763 Novel hum
310	19	100.0	195	3	AGG19885	Agg19885 Arabidops	383	19	100.0	222	3	AAU57061	Aau57061 Human pro
311	19	100.0	196	3	AGG00168	AgG00168 Human sec	384	19	100.0	222	3	AAU57061	Aau57061 Human pro
312	19	100.0	196	4	ABG11065	AbG11065 Novel hum	385	19	100.0	222	6	ADA36818	Ada36818 Acinetoba
313	19	100.0	196	4	ABG11065	AbG11065 Novel hum	386	19	100.0	222	7	ADD27209	Add27209 Human adi
314	19	100.0	197	5	AB54303	Ab54303 Lactococc	387	19	100.0	223	7	ADD27209	Add27209 Human adi
315	19	100.0	198	3	AB54303	Ab54303 Lactococc	388	19	100.0	224	3	AAU91479	Aay91479 Human sec
316	19	100.0	199	2	AAU91479	Aay91479 Mouse act	389	19	100.0	224	3	AAU91479	Aay91479 Human sec
317	19	100.0	199	2	AAU91479	Aay91479 Mouse act	390	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	391	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	392	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	393	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	394	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	395	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	396	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	397	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	398	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	399	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	400	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	401	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	402	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	403	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	404	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	405	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	406	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	407	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	408	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	409	19	100.0	224	3	AAU91479	Aay91479 Human sec
	19	100.0	199	2	AAU91479	Aay91479 Mouse act	410	19	100.0				

391	19	100.0	224	6	ADA54376	Ada54376 Human pro	464	19	100.0	257	2	AA977285	Aa977285 Human 26S
392	19	100.0	224	6	ADA56787	Ada56787 Human sec	465	19	100.0	257	2	AAW21729	Aaw21729 Nuclear m
393	19	100.0	224	7	ADC74046	Adc74046 Human sec	466	19	100.0	257	3	AA959110	Aag59110 Arabidops
394	19	100.0	224	7	ADC77843	Adc77843 Human sec	467	19	100.0	257	4	AA974447	Aag74447 Human col
395	19	100.0	226	4	AAW79581	Aam79581 Human pro	468	19	100.0	257	7	AA929888	Aag29888 C glutami
396	19	100.0	226	4	ABG03398	Abg03398 Novel hum	469	19	100.0	257	7	ADD27694	Add27694 Human adi
397	19	100.0	226	4	ABG09001	Abg09001 Novel hum	470	19	100.0	257	7	ADD26681	Add26681 Human adi
398	19	100.0	226	6	AA36076	Aa36076 Acinetoba	471	19	100.0	257	7	ADD27489	Add27489 Human adi
399	19	100.0	227	5	AA36076	Aa36076 Acinetoba	472	19	100.0	258	3	AA16704	Aab16704 Bacteriop
400	19	100.0	228	3	AAW74793	Aay74793 Neisseria	473	19	100.0	258	4	AAU36296	Aau36296 Pseudomon
401	19	100.0	228	4	AB47282	Ab47282 MSF-3752	474	19	100.0	258	6	ABU38502	Abu38502 Protein e
402	19	100.0	228	6	ABP78051	Abp78051 N. gonorr	475	19	100.0	260	4	AAW78433	Aam78433 Human pro
403	19	100.0	228	6	ABF80081	Abf80081 N. gonorr	476	19	100.0	260	4	AAU58456	Aau58456 Propionib
404	19	100.0	228	6	ABG75884	Abg75884 Mycobacte	477	19	100.0	260	4	ABG30334	Abg30334 Novel hum
405	19	100.0	229	4	AA81949	Aag81949 S. epider	478	19	100.0	260	6	ABM54975	Abm54975 Propionib
406	19	100.0	229	4	AAU04880	Aau04880 Micromono	479	19	100.0	261	4	AAU00701	Aau00701 Mouse FCT
407	19	100.0	229	5	AB931524	Ab931524 Herbicida	480	19	100.0	261	5	ABB79646	Abb79646 Mouse FCT
408	19	100.0	230	4	AA800078	Aab80078 Corynebac	481	19	100.0	261	5	ABG78505	Abg78505 Murine FC
409	19	100.0	230	4	AG90273	Aag90273 C glutami	482	19	100.0	261	5	AAU82959	Aau82959 Human hom
410	19	100.0	231	2	AAW60388	Aaw60388 Streptoco	483	19	100.0	261	6	ABU41453	Abu41453 Protein e
411	19	100.0	231	5	ABM50143	Abm50143 Listeria	484	19	100.0	261	7	ADE54842	Ade54842 Rat Prote
412	19	100.0	231	6	ADA36637	Ada36637 Acinetoba	485	19	100.0	261	7	ABG54844	Abg54844 Human Pro
413	19	100.0	232	2	AA27368	Aay27368 Group B S	486	19	100.0	262	4	ABG04663	Abg04663 Novel hum
414	19	100.0	232	3	AA11258	Aag11258 Arabidops	487	19	100.0	262	5	AAU72572	Aau72572 Arabidops
415	19	100.0	234	3	AA322116	Aab322116 Maize glu	488	19	100.0	263	2	AAW95344	Aaw95344 Human adu
416	19	100.0	234	3	AA326676	Aag26676 Arabidops	489	19	100.0	263	4	AA905050	Aab905050 Human sec
417	19	100.0	234	5	AB911390	Ab911390 Herbicida	490	19	100.0	263	5	ABG54646	Abg54646 Human alb
418	19	100.0	234	7	ADC96896	Adc96896 E. faeciu	491	19	100.0	263	7	ADA44959	Ada44959 Human pol
419	19	100.0	236	3	AA342087	Aab42087 Human ORF	492	19	100.0	264	3	AA15729	Aag15729 Arabidops
420	19	100.0	236	4	AAU36145	Aau36145 Klebsiell	493	19	100.0	264	4	AA86363	Aab86363 Human cer
421	19	100.0	236	4	AB61537	Ab61537 DegP prot	494	19	100.0	264	4	ABG18250	Abg18250 Novel hum
422	19	100.0	236	4	ABG27447	Abg27447 Novel hum	495	19	100.0	265	2	AAO8269	Aao8269 Mouse ser
423	19	100.0	238	3	AA11695	Aab11695 Mouse ser	496	19	100.0	265	2	AAW70534	Aaw70534 Mouse neu
424	19	100.0	238	3	AB328132	Aab328132 Glyceral	497	19	100.0	265	6	ABU24490	Abu24490 Protein e
425	19	100.0	238	4	AA61538	Ab61538 DegP prot	498	19	100.0	266	4	AAW79417	Aaw79417 Human pro
426	19	100.0	238	4	AG83029	Aag83029 S. epider	499	19	100.0	267	2	AAW85606	Aaw85606 Secreted
427	19	100.0	238	4	AA83016	Aag83016 S. epider	500	19	100.0	267	2	ABU50254	Abu50254 Protein e
428	19	100.0	238	5	AB53557	Ab53557 Lactococc	501	19	100.0	268	5	AAE15256	Aae15256 Human RNA
429	19	100.0	238	5	AAU0750	Aau0750 Eucalyptu	502	19	100.0	268	5	ABP61845	Abp61845 Human pol
430	19	100.0	238	6	ABU44475	Abu44475 Protein e	503	19	100.0	269	2	AAW55786	Aaw55786 Human aqu
431	19	100.0	238	6	ABU42738	Abu42738 Protein e	504	19	100.0	269	2	AAW94322	Aaw94322 Human aqu
432	19	100.0	238	6	ABW72778	Abm72778 Staphyloc	505	19	100.0	269	4	ABG26217	Abg26217 Novel hum
433	19	100.0	239	3	AAW74795	Aay74795 Neisseria	506	19	100.0	269	6	ABR53586	AbR53586 Protein s
434	19	100.0	239	3	AAW74794	Aay74794 Neisseria	507	19	100.0	269	6	ABP58050	Abp58050 Human aqu
435	19	100.0	239	3	AA11257	Aag11257 Arabidops	508	19	100.0	269	7	ABE64028	AbE64028 Human pro
436	19	100.0	239	4	AB96778	Aab96778 Putative	509	19	100.0	270	4	AA90189	Aag90189 C glutami
437	19	100.0	240	3	AA19884	Aag19884 Arabidops	510	19	100.0	270	6	ADA05788	Ada05788 Human NOV
438	19	100.0	240	3	AA24297	Aag24297 Arabidops	511	19	100.0	270	7	ADC08133	Adc08133 Rice prot
439	19	100.0	240	7	AB25064	Ab25064 Plant gro	512	19	100.0	271	2	AA15227	Aay15227 Human rec
440	19	100.0	241	5	ABP40753	Abp40753 Staphyloc	513	19	100.0	271	3	AA28205	Ab28205 Novel hum
441	19	100.0	241	6	ABU28310	Abu28310 Protein e	514	19	100.0	271	3	AA31287	Aag31287 Arabidops
442	19	100.0	243	2	AAW56089	Aay56089 S. pneumo	515	19	100.0	271	3	AA22050	Aag22050 Arabidops
443	19	100.0	243	5	ABP38742	Abp38742 Staphyloc	516	19	100.0	271	4	AA92851	Aab92851 Human pro
444	19	100.0	244	5	AB883337	Ab883337 Human TSA	517	19	100.0	271	4	AA888446	Aab88446 Human mem
445	19	100.0	244	6	ABU44263	Abu44263 Protein e	518	19	100.0	271	4	AA64357	Aag64357 Human sig
446	19	100.0	245	4	ABG10688	Abg10688 Novel hum	519	19	100.0	271	5	AB90282	Abb90282 Human pol
447	19	100.0	245	7	ADC08337	Adc08337 Rice prot	520	19	100.0	271	7	AB95444	Ab95444 E. faeciu
448	19	100.0	247	7	ABE66966	AbE66966 Human pan	521	19	100.0	272	4	AAU23799	Aau23799 Novel hum
449	19	100.0	248	2	AA11160	Aay11160 S. pneumo	522	19	100.0	272	4	AA94302	Aam94302 Human rep
450	19	100.0	249	4	AA11160	Aay11160 S. pneumo	523	19	100.0	272	4	ABG16020	Abg16020 Novel hum
451	19	100.0	249	4	ABU19896	Abu19896 Polyketid	524	19	100.0	272	6	ABU42431	Abu42431 Protein e
452	19	100.0	249	6	ABU11227	Abu11227 Human TAN	525	19	100.0	272	6	ABM71001	Abm71001 Staphyloc
453	19	100.0	249	7	AB861593	Ab861593 Z. mays a	526	19	100.0	273	3	ABM11696	Abm11696 Mouse ser
454	19	100.0	250	2	AA60153	Aar60153 Nematode-	527	19	100.0	273	3	AA15728	Aag15728 Arabidops
455	19	100.0	250	2	AA33913	AA33913 Sequence	528	19	100.0	273	4	AA86365	Ab86365 Human cer
456	19	100.0	250	2	AA77082	Aar77082 Tobacco r	529	19	100.0	273	6	ABR42553	ABR42553 Simocycli
457	19	100.0	250	3	AA842541	Aab42541 Human ORF	530	19	100.0	273	7	ABE08633	Abe08633 Novel pro
458	19	100.0	250	4	AAU47372	Aau47372 Propionib	531	19	100.0	274	3	AA800202	Aab00202 Putative
459	19	100.0	250	6	ABM43891	Abm43891 Propionib	532	19	100.0	274	3	AA800211	Aab00211 Potential
460	19	100.0	253	4	AAU34686	Aau34686 E. Coli c	533	19	100.0	274	4	AA633322	Aab63322 Human bre
461	19	100.0	253	7	ADC95445	Adc95445 E. faeciu	534	19	100.0	274	6	ABU23635	Abu23635 Protein e
462	19	100.0	254	6	ADA05790	Ada05790 Human NOV	535	19	100.0	274	6	ABU46447	Abu46447 Protein e
463	19	100.0	256	6	ABM67840	Abm67840 Photorhab	536	19	100.0	274	7	ADC32869	Adc32869 Human nov

537	19	100.0	275	6	ABU241121	Abu24121 Protein e	610	19	100.0	300	7	ADC31643	Adc31643 Human nov
538	19	100.0	276	3	AAB42611	Aab42611 Human ORF	611	19	100.0	301	4	AAG82151	Aag82151 S. epider
539	19	100.0	277	2	AAW18875	Aaw18875 Babesia e	612	19	100.0	302	3	AAG15595	Aag15595 Arabidops
540	19	100.0	278	3	AAG22852	Aag22852 Arabidops	613	19	100.0	303	4	AAG20322	Aag20322 Arabidops
541	19	100.0	279	3	AAG04958	Aag04958 Arabidops	614	19	100.0	303	4	AAG89980	Aag89980 C. Glutami
542	19	100.0	278	3	AAG07719	Aag07719 Arabidops	615	19	100.0	304	3	AAG37633	Aag37633 Arabidops
543	19	100.0	278	3	AAG39551	Aag39551 Arabidops	616	19	100.0	304	3	AAG04957	Aag04957 Arabidops
544	19	100.0	278	3	AAG37634	Aag37634 Arabidops	617	19	100.0	304	3	AAG30063	Aag30063 Arabidops
545	19	100.0	278	5	ABP30329	Abp30329 Streptoco	618	19	100.0	304	3	AAG07718	Aag07718 Arabidops
546	19	100.0	278	6	ABP77659	Abp77659 N. gonorr	619	19	100.0	304	3	AAG39550	Aag39550 Arabidops
547	19	100.0	278	6	ABU37537	Abu37537 Protein e	620	19	100.0	305	3	AAG30713	Aag30713 Arabidops
548	19	100.0	278	6	ABU37837	Abu37837 Protein e	621	19	100.0	305	3	AAG10620	Aag10620 Arabidops
549	19	100.0	279	3	AAG54754	Aag54754 Arabidops	622	19	100.0	305	4	AAG92837	Aag92837 C. Glutami
550	19	100.0	279	5	ABG30336	Abg30336 Novel hum	623	19	100.0	306	3	AAG40089	Aag40089 Arabidops
551	19	100.0	279	5	ABG55091	Abg55091 Lactococc	624	19	100.0	306	3	AAG13663	Aag13663 Arabidops
552	19	100.0	280	2	AAW03521	Aaw03521 Bali modi	625	19	100.0	306	4	AAU03620	Aau03620 Group B S
553	19	100.0	280	2	AAW37073	Aaw37073 Amino aci	626	19	100.0	306	5	ABP29888	Abp29888 Streptoco
554	19	100.0	280	4	AAG90751	Aag90751 C. Glutami	627	19	100.0	306	5	ABP28497	Abp28497 Streptoco
555	19	100.0	281	3	AAG26675	Aag26675 Arabidops	628	19	100.0	306	7	ADP70040	Adp70040 C. neofor
556	19	100.0	281	5	ABG77963	Abg77963 Ehrlichia	629	19	100.0	307	7	ADC37261	Adc37261 Nuclear f
557	19	100.0	281	6	ADA09791	Ada09791 E. canis	630	19	100.0	308	3	AAE53390	Aae53390 Human col
558	19	100.0	282	2	AAW34205	Aaw34205 Streptomy	631	19	100.0	308	4	AAG52974	Aag52974 C. Glutami
559	19	100.0	282	2	AAW55806	Aaw55806 Streptomy	632	19	100.0	308	5	ABE54256	Abbe54256 Lactococc
560	19	100.0	282	3	AAW57116	Aaw57116 Human pro	633	19	100.0	308	5	ABP60904	Abp60904 Lactococc
561	19	100.0	282	3	AAG13664	Aag13664 Arabidops	634	19	100.0	308	5	ABP65647	Abp65647 Bifidobac
562	19	100.0	282	3	AAG40090	Aag40090 Arabidops	635	19	100.0	308	7	ADP79410	Adp79410 Human G p
563	19	100.0	282	6	ABM67054	Abm67054 Photorhab	636	19	100.0	309	3	AAG31285	Aag31285 Arabidops
564	19	100.0	282	6	ABU19351	Abu19351 Protein e	637	19	100.0	309	5	ABP93102	Abp93102 Herbicida
565	19	100.0	283	2	AAW89755	Aaw89755 Staphyloc	638	19	100.0	309	5	ABP61176	Abp61176 Human GFC
566	19	100.0	284	4	AAB93850	Aab93850 Human pro	639	19	100.0	310	3	AAE21021	Aae21021 Human nuc
567	19	100.0	284	2	AAW95957	Aaw95957 Eukaryoti	640	19	100.0	310	5	AAW51006	Aaw51006 Corynebac
568	19	100.0	285	6	ABU21606	Abu21606 Protein e	641	19	100.0	310	6	AAG79924	Aag79924 Human Sir
569	19	100.0	285	6	ABU41998	Abu41998 Protein e	642	19	100.0	310	6	ABU39481	Abu39481 Protein e
570	19	100.0	286	6	ADA05784	Ada05784 Human NOV	643	19	100.0	311	3	AAE11697	Aae11697 Mouse eer
571	19	100.0	286	6	ABU41392	Abu41392 Protein e	644	19	100.0	311	3	AAE46265	Aae46265 Arabidops
572	19	100.0	287	6	AAG26674	Aag26674 Arabidops	645	19	100.0	311	4	AAE11899	Aae11899 Human G-p
573	19	100.0	287	6	ABU00416	Abu00416 Human nov	646	19	100.0	311	4	AAG71742	Aag71742 Human OR-
574	19	100.0	288	6	ABM69140	Abm69140 Photorhab	647	19	100.0	311	4	AAG72431	Aag72431 Human OR-
575	19	100.0	288	6	AAW38988	Aaw38988 Protein e	648	19	100.0	311	4	AAG71759	Aag71759 Human olf
576	19	100.0	289	6	AAW35600	Aaw35600 Haemophil	649	19	100.0	311	4	AAE10690	Aae10690 G-protein
577	19	100.0	289	6	ABU30505	Abu30505 Protein e	650	19	100.0	311	4	AAG89917	Aag89917 C. Glutami
578	19	100.0	290	3	AAW78511	Aaw78511 Human unc	651	19	100.0	311	4	AAU24520	Aau24520 Human olf
579	19	100.0	290	3	AAG23887	Aag23887 Arabidops	652	19	100.0	311	5	ABP95725	Abp95725 Human GFC
580	19	100.0	291	6	ADA05786	Ada05786 Human NOV	653	19	100.0	311	5	ABP95777	Abp95777 Human GFC
581	19	100.0	291	6	ABW90394	Abw90394 Human pol	654	19	100.0	311	5	AAU95637	Aau95637 Human olf
582	19	100.0	291	7	ADC37617	Adc37617 Human nov	655	19	100.0	311	5	AAU95535	Aau95535 Human olf
583	19	100.0	291	7	ADC37617	Adc37617 Human GPC	656	19	100.0	311	5	ABJ04727	Abj04727 GPCR 7 pr
584	19	100.0	292	3	AAG16171	Aag16171 Arabidops	657	19	100.0	311	5	AAE71188	Aae71188 Human GPC
585	19	100.0	292	3	AAG60250	Aag60250 Arabidops	658	19	100.0	311	5	AAU85144	Aau85144 G-coupled
586	19	100.0	293	6	ABU48542	Abu48542 Protein e	659	19	100.0	311	5	AAU11103	Aau11103 Human nov
587	19	100.0	294	5	ABG91459	Abg91459 Purine/py	660	19	100.0	311	6	ABU01429	Abu01429 S. pneumo
588	19	100.0	294	6	ABR53328	AbR53328 Protein s	661	19	100.0	311	6	ABU11179	Abu11179 Human G-p
589	19	100.0	294	6	ABP98029	Abp98029 Protein p	662	19	100.0	311	6	ABU11195	Abu11195 Human G-p
590	19	100.0	295	3	AAG31286	Aag31286 Arabidops	663	19	100.0	311	7	ADC85801	Adc85801 Human GPC
591	19	100.0	295	5	ABP28274	Abp28274 Streptoco	664	19	100.0	311	7	ADC85977	Adc85977 Human GPC
592	19	100.0	295	5	ABG91460	Abg91460 Purine/py	665	19	100.0	311	7	ADD12777	Add12777 Novel hum
593	19	100.0	295	5	ABG91461	Abg91461 Purine/py	666	19	100.0	312	3	AAG39549	Aag39549 Arabidops
594	19	100.0	295	5	ABG91449	Abg91449 Purine/py	667	19	100.0	312	3	AAG04956	Aag04956 Arabidops
595	19	100.0	296	5	AAW52345	Aaw52345 Aromatic	668	19	100.0	312	3	AAG07717	Aag07717 Arabidops
596	19	100.0	297	6	ABU22319	Abu22319 Protein e	669	19	100.0	312	3	AAG37632	Aag37632 Arabidops
597	19	100.0	297	6	ABU32270	Abu32270 Protein e	670	19	100.0	312	4	AAE80162	Aae80162 Corynebac
598	19	100.0	299	2	AAW06968	Aaw06968 E. canis	671	19	100.0	312	4	AAE72585	Aae72585 Human OR-
599	19	100.0	299	3	AAW07975	Aaw07975 A corn tr	672	19	100.0	312	4	AAG71811	Aag71811 Human G-p
600	19	100.0	299	4	AAW93403	Aaw93403 Human pro	673	19	100.0	312	4	AAE06767	Aae06767 C. Glutami
601	19	100.0	299	4	AAW87731	Aaw87731 Human t2R	674	19	100.0	312	4	AAW91253	Aaw91253 Human olf
602	19	100.0	299	5	ABG77957	Abg77957 Ehrlichia	675	19	100.0	312	4	AAU24609	Aau24609 Human olf
603	19	100.0	299	6	ABR38899	AbR38899 Human tR2	676	19	100.0	312	5	ABW06624	Abw06624 G protein
604	19	100.0	299	6	ADA09779	Ada09779 E. canis	677	19	100.0	312	5	ABW06623	Abw06623 G protein
605	19	100.0	299	7	ADC97734	Adc97734 Rhizobito	678	19	100.0	312	5	ABP95923	Abp95923 Human GPC
606	19	100.0	300	3	AAG45274	Aag45274 Arabidops	679	19	100.0	312	5	AAU95531	Aau95531 Human olf
607	19	100.0	300	3	AAG30714	Aag30714 Arabidops	680	19	100.0	312	5	AAU85229	Aau85229 G-coupled
608	19	100.0	300	3	AAG23886	Aag23886 Arabidops	681	19	100.0	312	6	ABR52886	AbR52886 Protein s
609	19	100.0	300	5	ABP27828	Abp27828 Streptoco	682	19	100.0	312	6	ABU26073	Abu26073 Protein e

583	19	100.0	312	7	ADC65589	ADC85589 Human GPC	756	19	100.0	328	4	AAU35342	Aau35342 Enterococ
584	19	100.0	312	7	ADBI3513	Adi3513 C. Glutam	757	19	100.0	328	5	ABB91780	Abb91780 Herbicida
585	19	100.0	313	2	AAK37346	Aar37346 PEP PM. 1	758	19	100.0	328	6	ABO23285	Abc23285 Mouse ary
586	19	100.0	313	5	ABBA9170	Abb9170 Listeria	759	19	100.0	328	6	ABO23284	Abc23284 Cow aryl-
587	19	100.0	314	3	AAV99818	Aay99818 Wheat par	760	19	100.0	328	6	ABU28954	Abu28954 Protein e
588	19	100.0	314	6	ABRO1619	Abro1619 Human G p	761	19	100.0	328	6	ABM73204	Abm73204 Staphyloc
589	19	100.0	314	6	ABRO1585	Abro1585 Human G p	762	19	100.0	329	2	AAW80623	Aaw80623 S. pneumo
590	19	100.0	314	6	ABU17461	Abu17461 Protein e	763	19	100.0	329	5	ABP27977	Abp27977 Streptoco
591	19	100.0	315	4	ABBS2707	Abb52707 Escherich	764	19	100.0	329	7	ABM73892	Abm73892 DNA clone
592	19	100.0	315	4	AAV71585	Aag71585 Human c1f	765	19	100.0	331	2	AAV37504	Aay37504 Protein i
593	19	100.0	315	4	AAV72285	Aag72285 Human c1f	766	19	100.0	331	3	AAV43812	Aag43812 Arabidops
594	19	100.0	315	4	AAV72890	Aag72890 Human c1f	767	19	100.0	331	4	AAU36168	Aag36168 Klebsiell
595	19	100.0	315	5	AAE17492	Aae17492 Human sec	768	19	100.0	331	4	AAV72652	Aag72652 Murine OR
596	19	100.0	315	5	AAV71213	Aab71213 Human GPC	769	19	100.0	334	2	AAV36939	Aay36939 Chlamydia
597	19	100.0	315	6	ABM70746	Abm70746 Photorhab	770	19	100.0	334	2	AAE06056	Aae06056 Human gen
598	19	100.0	315	6	ABM70746	Abm70746 Photorhab	771	19	100.0	334	5	ABP41314	Abp41314 Human ova
599	19	100.0	315	7	ADC37263	Adc37263 Nuclear f	772	19	100.0	334	5	ABG33878	Abg33878 Human sec
600	19	100.0	315	7	ADD12827	Add12827 Novel hum	773	19	100.0	334	6	ADA57046	Ada57046 Human sec
701	19	100.0	315	7	ADE90061	Ade90061 Human rho	774	19	100.0	334	6	ADA40900	Ada40900 Human sec
702	19	100.0	316	3	AAV20321	Aag20321 Arabidops	775	19	100.0	334	6	ABU43221	Abu43221 Protein e
703	19	100.0	316	4	AAU55622	Aau55622 Propionib	776	19	100.0	334	6	ABR61839	Abri61839 Arabidops
704	19	100.0	316	6	ABP98031	Abp98031 Protein p	777	19	100.0	335	2	AAV57736	Aar57736 Hyperseus
705	19	100.0	316	6	ABM52141	Abm52141 Propionib	778	19	100.0	335	3	AAV87079	Aay87079 Human sec
706	19	100.0	317	2	AAV01381	Aay01381 Rat osteo	779	19	100.0	336	3	AAV46263	Aag46263 Arabidops
707	19	100.0	317	3	ABM41662	Abm41662 Human ORF	780	19	100.0	336	3	AAV32249	Aag32249 Arabidops
708	19	100.0	317	3	ABM69850	Abm69850 Photorhab	781	19	100.0	336	3	AAV11256	Aag11256 Arabidops
709	19	100.0	317	7	ADE57454	Ade57454 Rat Prote	782	19	100.0	336	6	ABU50329	Abu50329 Protein e
710	19	100.0	318	2	AAW21746	Aaw21746 E3330-bin	783	19	100.0	336	6	ABU50329	Abu50329 Protein e
711	19	100.0	318	2	AAW48894	Aaw48894 Homo sapi	784	19	100.0	337	3	AAV67495	Aay67495 Arabidops
712	19	100.0	318	2	AAW52864	Aaw52864 Apurinic/	785	19	100.0	337	3	AAV43811	Aag43811 Arabidops
713	19	100.0	318	2	AAW52864	Aaw52864 Apurinic/	786	19	100.0	337	3	AAV06421	Aag06421 Arabidops
714	19	100.0	318	3	AAV08737	Aag08737 Arabidops	787	19	100.0	337	3	AAV40549	Aag40549 Arabidops
715	19	100.0	318	3	AAV34921	Aag34921 Arabidops	788	19	100.0	338	6	ADA35864	Ada35864 Acinetoba
716	19	100.0	318	5	ABP60935	Abp60935 Zymomonas	789	19	100.0	339	2	AAV07063	Aay07063 Renal can
717	19	100.0	318	5	ADDO8904	Ado8904 Apurinic/	790	19	100.0	339	4	AAV61273	Aab61273 Human TAN
718	19	100.0	318	7	ADE25727	Ade25727 Human pro	791	19	100.0	339	4	AAV61274	Aab61274 Human TAN
719	19	100.0	319	3	AAV31049	Aag31049 Arabidops	792	19	100.0	339	4	AAV61275	Aab61275 Human TAN
720	19	100.0	319	3	ABM61257	Abm61257 Mature hu	793	19	100.0	339	4	AAV61276	Aab61276 Human TAN
721	19	100.0	319	4	ABM49403	Abm49403 Human gly	794	19	100.0	339	4	AAV72790	Aay72790 Human pla
722	19	100.0	319	6	ABP98028	Abp98028 Protein p	795	19	100.0	339	4	AAV38760	Aam38760 Human pol
723	19	100.0	319	6	ABU11223	Abu11223 Human TAN	796	19	100.0	339	4	AAV38760	Aam38760 Human pol
724	19	100.0	320	3	AAV50998	Aag50998 Arabidops	797	19	100.0	339	4	AAV38760	Aam38760 Human pol
725	19	100.0	320	3	AAV06286	Aag06286 Arabidops	798	19	100.0	339	4	AAV95490	Aab95490 Human pro
726	19	100.0	320	3	AAV40088	Aag40088 Arabidops	799	19	100.0	339	6	ABP72608	Abp72608 Benign pr
727	19	100.0	320	3	AAV13662	Aag13662 Arabidops	800	19	100.0	339	6	ABP72607	Abp72607 Benign pr
728	19	100.0	320	4	ABG82947	Abg82947 S. epider	801	19	100.0	339	6	ADA84105	Ada84105 Human GP6
729	19	100.0	320	5	ABG91676	Abg91676 Herbicida	802	19	100.0	339	6	ABU11221	Abu11221 Human TAN
730	19	100.0	321	3	AAV99817	Aay99817 Soybean p	803	19	100.0	339	6	ABU11239	Abu11239 Glycoprot
731	19	100.0	321	5	ABP51575	Abp51575 Human G p	804	19	100.0	339	6	ABU11240	Abu11240 Glycoprot
732	19	100.0	321	6	ADA89479	Ada89479 Staphyloc	805	19	100.0	339	6	ABU11241	Abu11241 Glycoprot
733	19	100.0	322	3	AAV94666	Aay94666 Human unc	806	19	100.0	339	6	ABU11242	Abu11242 Glycoprot
734	19	100.0	322	3	AAV94666	Aay94666 Murine un	807	19	100.0	339	6	ABU70998	Abu70998 Human adi
735	19	100.0	322	3	AAV20529	Aag20529 Arabidops	808	19	100.0	340	6	ABR42679	Abri42679 Pseudomon
736	19	100.0	322	3	AAV08879	Aag08879 Arabidops	809	19	100.0	341	4	ABG5957	Abg5957 Novel hum
737	19	100.0	322	6	ABU25231	Abu25231 Protein e	810	19	100.0	342	4	ABG52472	Abg52472 Escherich
738	19	100.0	322	6	ABM72033	Abm72033 Staphyloc	811	19	100.0	342	4	ABM11919	Abm11919 Human cyt
739	19	100.0	323	3	AAV06285	Aag06285 Arabidops	812	19	100.0	342	4	ABM11919	Abm11919 Human cyt
740	19	100.0	323	6	ABU41982	Abu41982 Protein e	813	19	100.0	342	4	AAV79442	Aam79442 Human pro
741	19	100.0	324	3	AAV06284	Aag06284 Arabidops	814	19	100.0	342	5	ABG79682	Abg79682 Tumour in
742	19	100.0	324	6	ABU33384	Abu33384 Protein e	815	19	100.0	343	3	AAV15727	Aag15727 Arabidops
743	19	100.0	324	6	ABU29258	Abu29258 Protein e	816	19	100.0	343	3	AAV45273	Aag45273 Arabidops
744	19	100.0	324	6	ADA33349	Ada33349 Acinetoba	817	19	100.0	343	5	ABG22851	Abg22851 Herbicida
745	19	100.0	325	2	AAV50034	Aay50034 Porcine c	818	19	100.0	344	3	AAV22851	Aag22851 Arabidops
746	19	100.0	325	3	AAV94669	Aay94669 Murine un	819	19	100.0	345	3	AAV56610	Aab56610 Human pro
747	19	100.0	325	3	AAV94665	Aay94665 Human unc	820	19	100.0	345	7	ADC94222	Adc94222 E. faeciu
748	19	100.0	325	3	AAV20320	Aag20320 Arabidops	821	19	100.0	346	3	AAV32505	Aab32505 S. lavend
749	19	100.0	325	6	ADA05792	Ada05792 Human NOV	822	19	100.0	346	4	AAV40546	Aam40546 Human pol
750	19	100.0	326	3	AAV23939	Aab23939 Hepatitis	823	19	100.0	346	7	ADE10277	Ade10277 S. lavend
751	19	100.0	326	3	AAV46264	Aag46264 Arabidops	824	19	100.0	347	4	AAV93168	Aag93168 C. glurami
752	19	100.0	326	3	AAV51053	Aag51053 Arabidops	825	19	100.0	347	5	ABF38453	Abf38453 Staphyloc
753	19	100.0	326	5	ABG93190	Abg93190 Herbicida	826	19	100.0	347	5	ABP41750	Abp41750 Human ova
754	19	100.0	327	2	AAV50033	Aay50033 Porcine c	827	19	100.0	347	7	ADC54077	Adc54077 Mouse tes
755	19	100.0	327	6	ABU11347	Abu11347 Protein e	828	19	100.0	347	7	ADC54079	Adc54079 Mouse tes

829	19	100.0	350	3	AAB02086	Aab02086 F420-depe	902	19	100.0	377	6	ABU25867	Abu25867 Protein e
830	19	100.0	350	5	AAO15202	Aao15202 Rhodococc	903	19	100.0	378	4	AAm93361	Aam93361 Human pol
831	19	100.0	350	5	Aau97147	Aau97147 Rhodococc	904	19	100.0	378	4	AAe20087	Aae20087 Lactobaci
832	19	100.0	350	5	AAE16748	Aae16748 Rhodococc	905	19	100.0	379	6	ABU23947	Abu23947 Protein e
833	19	100.0	351	6	ABU18964	Abu18964 Protein e	906	19	100.0	379	6	ABU49216	Abu49216 Protein e
834	19	100.0	352	6	ABb48732	Abb48732 Listeria	907	19	100.0	379	6	ABM04817	Abm04817 Rat cdc 3
835	19	100.0	352	6	ABU32749	Abu32749 Protein e	908	19	100.0	379	7	ADD45256	ADD45256 H. pylori
836	19	100.0	353	3	AAy94667	Aay94667 Human unc	909	19	100.0	380	2	AAW98800	Aaw98800 R. prote
837	19	100.0	353	3	AAg08899	Aag08899 Arabidops	910	19	100.0	380	4	AAU35698	Aau35698 Helicobac
838	19	100.0	353	4	ABe16162	Abel16162 Novel hum	911	19	100.0	380	5	ABG30683	Abg30683 Cystathio
839	19	100.0	354	3	AAg43810	Aag43810 Arabidops	912	19	100.0	380	6	ABU33267	Abu33267 Protein e
840	19	100.0	355	4	AAg74072	Aag74072 Human col	913	19	100.0	382	5	ABP43943	Abp43943 Unidentif
841	19	100.0	356	3	AAy73382	Aay73382 HTRM clon	914	19	100.0	382	6	ABU33959	Abu33959 Protein e
842	19	100.0	356	6	ADA36384	Ada36384 Acinetoba	915	19	100.0	383	4	ABE60853	Abe60853 Mevalonat
843	19	100.0	357	3	AAg29807	Aag29807 Arabidops	916	19	100.0	383	6	ABU15985	Abu15985 Protein e
844	19	100.0	357	3	AAg82732	Aag82732 S. epider	917	19	100.0	383	6	ABU97187	Abu97187 Enzyme po
845	19	100.0	357	5	ABP63001	Abp63001 Human pol	918	19	100.0	384	3	AAV71056	Aav71056 Human mem
846	19	100.0	357	5	ABb48768	Abb48768 Listeria	919	19	100.0	384	4	AAm40198	Aam40198 Human pol
847	19	100.0	358	3	AAg29806	Aag29806 Arabidops	920	19	100.0	384	4	AAm40199	Aam40199 Human pol
848	19	100.0	358	3	AAg15679	Aag15679 Arabidops	921	19	100.0	384	4	AAm40199	Aam40199 Human pol
849	19	100.0	359	3	AAg32423	Aag32423 Arabidops	922	19	100.0	384	4	AAm40199	Aam40199 Human pol
850	19	100.0	359	3	AAg48953	Aag48953 Arabidops	923	19	100.0	384	5	ABE54918	Abe54918 Lactococc
851	19	100.0	359	5	ABP74056	Abp74056 Candida a	924	19	100.0	384	6	ABO23281	AbO23281 Human ary
852	19	100.0	360	4	AAm40979	Aam40979 Human pol	925	19	100.0	384	6	ABO23282	AbO23282 Chimpanze
853	19	100.0	360	7	ADe09991	Ade09991 Novel pro	926	19	100.0	385	3	AAg32850	Aag32850 Arabidops
854	19	100.0	361	3	AAg08898	Aag08898 Arabidops	927	19	100.0	385	5	AAU72542	Aau72542 Arabidops
855	19	100.0	361	3	AAg45156	Aag45156 Arabidops	928	19	100.0	386	3	AAU13568	Aau13568 Streptomy
856	19	100.0	361	3	AAg08897	Aag08897 Arabidops	929	19	100.0	386	4	AAU35448	Aau35448 Haemophil
857	19	100.0	367	3	AAg45155	Aag45155 Arabidops	930	19	100.0	386	5	ABU30277	Abu30277 Protein e
858	19	100.0	367	3	AAO13149	Aao13149 Human pol	931	19	100.0	388	6	ABR55193	AbR55193 Amino aci
859	19	100.0	369	4	AAy72791	Aay72791 Human GPV	932	19	100.0	389	4	ABR55193	AbR55193 Amino aci
860	19	100.0	369	4	ABG01143	Abg01143 Novel hum	933	19	100.0	389	6	ABR55193	AbR55193 Amino aci
861	19	100.0	369	4	ABG13946	Abg13946 Novel hum	934	19	100.0	389	6	ABU97194	Abu97194 Enzyme po
862	19	100.0	369	5	ABP28223	Abp28223 Streptoco	935	19	100.0	389	6	ABU97193	Abu97193 Enzyme po
863	19	100.0	369	5	ABP27311	Abp27311 Streptoco	936	19	100.0	390	6	ABU10407	Abu10407 Artirobac
864	19	100.0	369	6	ABU01488	Abu01488 S. pneumo	937	19	100.0	390	6	ABU10412	Abu10412 Artirobac
865	19	100.0	369	6	ABU46555	Abu46555 Protein e	938	19	100.0	390	7	ABD74583	Abd74583 Mycobacte
866	19	100.0	369	6	ABU28974	Abu28974 Protein e	939	19	100.0	391	2	AAW37992	Aaw37992 Mutant As
867	19	100.0	369	6	ABU46032	Abu46032 Human nov	940	19	100.0	391	2	AAV39873	Aav39873 Mutant As
868	19	100.0	369	7	ADC33076	Adc33076 Mouse gro	941	19	100.0	391	5	ABG05608	Abg05608 Mutant As
869	19	100.0	370	3	AAb48663	Aab48663 Mouse gro	942	19	100.0	391	5	ABG05608	Abg05608 Mutant As
870	19	100.0	370	4	AAU00700	Aau00700 Mouse VEG	943	19	100.0	392	5	ABG05608	Abg05608 Mutant As
871	19	100.0	370	4	AAU00700	Aau00700 Mouse FCT	944	19	100.0	393	6	ADe89536	Ade89536 Staphyloc
872	19	100.0	370	4	AAg65602	Aag65602 Mouse zve	945	19	100.0	393	6	ADe89536	Ade89536 Staphyloc
873	19	100.0	370	5	ABP79645	Abp79645 Mouse FCT	946	19	100.0	394	5	ABG61582	Abg61582 High grow
874	19	100.0	370	5	ABP79589	Abp79589 Mouse zve	947	19	100.0	394	5	ABG61582	Abg61582 High grow
875	19	100.0	370	5	ABG78504	Abg78504 Murine FC	948	19	100.0	394	5	ABG61582	Abg61582 High grow
876	19	100.0	370	5	ABP1641	Abp1641 Mouse zve	949	19	100.0	394	5	ABG61582	Abg61582 High grow
877	19	100.0	370	6	ABU72440	Abu72440 Mouse gro	950	19	100.0	396	4	AAU69440	Aau69440 Human pur
878	19	100.0	370	6	ABR43945	Abr43945 Mouse PDG	951	19	100.0	396	4	AAU69440	Aau69440 Human pur
879	19	100.0	370	7	AAE38477	Aae38477 Mouse pla	952	19	100.0	396	4	AAU69440	Aau69440 Human pur
880	19	100.0	371	5	ABP27978	Abp27978 Streptoco	953	19	100.0	397	3	AAV81772	Aav81772 Streptoco
881	19	100.0	371	5	ABP29850	Abp29850 Streptoco	954	19	100.0	397	3	AAV81772	Aav81772 Streptoco
882	19	100.0	371	5	ABP53305	Abp53305 Lactococc	955	19	100.0	397	3	AAV81772	Aav81772 Streptoco
883	19	100.0	371	5	ABU45379	Abu45379 Protein e	956	19	100.0	397	3	AAV81772	Aav81772 Streptoco
884	19	100.0	371	6	ABU45760	Abu45760 Protein e	957	19	100.0	397	3	AAV81772	Aav81772 Streptoco
885	19	100.0	371	6	ABU44096	Abu44096 Protein e	958	19	100.0	398	2	AAW56004	Aaw56004 Trichomon
886	19	100.0	371	6	ABD09252	Abd09252 Alloiococ	959	19	100.0	398	2	AAW56004	Aaw56004 Trichomon
887	19	100.0	372	6	ABO23283	AbO23283 Olive bab	960	19	100.0	398	2	AAW56004	Aaw56004 Trichomon
888	19	100.0	372	6	ABO23287	AbO23287 Common sq	961	19	100.0	398	2	AAW56004	Aaw56004 Trichomon
889	19	100.0	373	4	AAm40200	Aam40200 Human pol	962	19	100.0	398	2	AAW56004	Aaw56004 Trichomon
890	19	100.0	373	4	AAg45154	Aag45154 Arabidops	963	19	100.0	398	2	AAW56004	Aaw56004 Trichomon
891	19	100.0	374	3	AAg34157	Aag34157 Zea mays	964	19	100.0	399	3	AAW56004	Aaw56004 Trichomon
892	19	100.0	374	3	AAU38099	Aau38099 Streptoco	965	19	100.0	400	2	AAW17831	Aaw17831 Human mev
893	19	100.0	374	4	AAU37854	Aau37854 Streptoco	966	19	100.0	400	2	AAW17831	Aaw17831 Human mev
894	19	100.0	374	4	AAU01046	Aau01046 CFE 49 pr	967	19	100.0	400	2	AAW17831	Aaw17831 Human mev
895	19	100.0	374	4	AAb48046	Aab48046 Signal tr	968	19	100.0	400	2	AAW17831	Aaw17831 Human mev
896	19	100.0	374	6	ABU02774	Abu02774 S. pneumo	969	19	100.0	401	3	AAE62736	Aae62736 Human Pro
897	19	100.0	375	4	AAU35335	Aau35335 Enterococ	970	19	100.0	401	3	AAE62736	Aae62736 Human Pro
898	19	100.0	375	4	AAU35412	Aau35412 Human pro	971	19	100.0	401	3	AAE62736	Aae62736 Human Pro
899	19	100.0	375	4	ADb07634	Adb07634 Alloiococ	972	19	100.0	402	3	AAE62736	Aae62736 Human Pro
900	19	100.0	376	3	AAU6430	Aau6430 Arabidops	973	19	100.0	402	3	AAE62736	Aae62736 Human Pro
901	19	100.0	376	5	AAU72514	Aau72514 Arabidops	974	19	100.0	402	3	AAE62736	Aae62736 Human Pro

975 19 100.0 403 3 AAG50762 Arabidops  
 976 19 100.0 403 3 AAG13524 Arabidops  
 977 19 100.0 403 4 ABE58218 Drosophil  
 978 19 100.0 404 3 AAG53518 Arabidops  
 979 19 100.0 404 3 AAG39960 Arabidops  
 980 19 100.0 404 6 ABU40766 Protein e  
 981 19 100.0 405 2 AAY14954 Amino aci  
 982 19 100.0 405 4 AAB83254 Rat FATP4  
 983 19 100.0 405 6 ABM69291 Photorhab  
 984 19 100.0 406 2 AAR21549 Human Try  
 985 19 100.0 406 6 ABU29329 Protein e  
 986 19 100.0 407 4 AAM39361 Human pol  
 987 19 100.0 407 5 ABP27964 Streptoco  
 988 19 100.0 407 5 ABP61471 Human NF-  
 989 19 100.0 407 6 ABU46908 Protein e  
 990 19 100.0 408 4 AAG91382 C glutami  
 991 19 100.0 409 5 AAE30494 Fruit fly  
 992 19 100.0 410 2 AAR54663 Transcrip  
 993 19 100.0 410 2 AAR89212 Transcrip  
 994 19 100.0 410 5 AAE30503 Fruit fly  
 995 19 100.0 411 4 AAU56508 Propionib  
 996 19 100.0 411 6 ABM53027 Propionib  
 997 19 100.0 411 6 ABU50525 Protein e  
 998 19 100.0 412 3 AAB58972 Breast an  
 999 19 100.0 412 3 AAB57091 Human pro  
 1000 19 100.0 412 6 ABU14941 Protein e

## ALIGNMENTS

RESULT 1  
 ID RAM47151 standard; peptide; 6 AA.  
 AC RAM47151;  
 DT 12-FEB-2002 (first entry)

DE S chrysomallus actinomycin biosynthesis protein acmC fragment #9.

Modular enzyme system; cyclic gene synthesis; repetitive coding sequence;  
 antibiotic; non-ribosomal peptide synthetase; NRPS; PKS;  
 polyketide synthase; actinomycin biosynthesis.

Streptomyces chrysomallus.  
 Synthetic.

WO200181564-A2.

01-NOV-2001.

25-APR-2001; 2001WO-DE001578.

26-APR-2000; 2000DE-01021267.

(ACTI-) ACTINODRUG PHARM GMBH.

Schauwecker F;

WPI; 2002-049276/06.  
 N-PSDB; ABA03345.

Preparing DNA encoding modular protein for e.g. producing new enzymes for  
 synthesis of polyketide antibiotics, comprises cyclic integration of  
 fragments into a vector.

Example 3; Page 54; 83pp; German.

The present invention relates to the preparation of DNA, in a circular  
 vector, that encodes one or more segments of a modular polypeptide. DNA  
 or DNA libraries produced this way are used to produce modular  
 polypeptides, particularly enzymes, which can be used to act on

CC substrates to produce compounds for therapeutic testing. Enzymes of  
 CC particular interest are those involved in non-ribosomal peptide synthesis  
 CC or polyketide synthesis, and compounds for testing are particularly  
 CC macrolide antibiotics, including penicillins, vancomycins or  
 CC erythromycins, but may also be modular receptors. The present sequence is  
 CC a fragment of a protein encoded by a Streptomyces chrysomallus  
 CC actinomycin biosynthesis gene which was used in a plasmid in the  
 CC exemplification of the invention

XX Sequence 6 AA;

Query Match 100.0%; Score 19; DB 5; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB 3 VAEF 6

## RESULT 2

AAB07871 ID AAB07871 standard; peptide; 7 AA.

XX AAB07871;

DT 14-NOV-2000 (first entry)

DE A beta-secretase inhibitor peptide.

Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 inhibitor.

Synthetic.

Key Location/Qualifiers

Modified-site 3 /note= "hydroxyethylene"

WO200047619-A2.

17-AUG-2000.

10-FEB-2000; 2000WO-US003819.

10-FEB-1999; 99US-0119571P.

15-JUN-1999; 99US-0139172P.

(ELAN-) ELAN PHARM INC.

Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 Sinha S, Tateuno G, Tung J, Wang S, Mcconlogue L;

WPI; 2000-533011/48.

Purified beta-secretase protein used in assays to discover inhibitors  
 which can be used for the treatment of amyloidogenic diseases e.g.  
 Alzheimer's disease.

Disclosure; Page 12; 121pp; English.

The specification describes a beta-secretase enzyme. The enzyme cleaves  
 beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 enzyme is therefore implicated in the production of amyloid plaque  
 components which accumulate in the brains of individuals afflicted with  
 Alzheimer's disease. Inhibitors of beta-secretase are administered to a  
 mammalian subject e.g. with Alzheimer's disease or Alzheimer's disease-  
 like pathology to test if they maintain or improve cognitive ability or  
 reduce the plaque burden. The compounds are used for the treatment of  
 amyloidogenic diseases e.g. Alzheimer's disease. The present sequence  
 represents an inhibitor of beta-secretase enzyme

Q Sequence 7 AA;

Query Match 100.0%; Score 19; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
|  
b 4 VAEF 7

RESULT 3

AB07872  
D AAB07872 standard; peptide; 8 AA.

X AAB07872;

X 14-NOV-2000 (first entry)

E A beta-secretase inhibitor peptide.

X Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
W amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
W inhibitor.

X Synthetic.

H Key Location/Qualifiers

T Modified-site 4 /note= "hydroxyethylene"

X WO200047618-A2.

X 17-AUG-2000.

F 10-FEB-2000; 2000WO-US003819.

X 10-FEB-1999; 99US-0119571P.

R 15-JUN-1999; 99US-0139172P.

X (ELAN-) ELAN PHARM INC.

X Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;

I Sinha S, Tatsuno G, Tung J, Wang S, Mcconlogue L;

X WPI; 2000-533011/48.

T Purified beta-secretase protein used in assays to discover inhibitors  
which can be used for the treatment of amyloidogenic diseases e.g.

T Alzheimer's disease.

X Disclosure; Page 12; 121pp; English.

X The specification describes a beta-secretase enzyme. The enzyme cleaves  
C beta-amyloid precursor protein to produce beta-amyloid peptide. This  
C enzyme is therefore implicated in the production of amyloid plaque  
C components which accumulate in the brains of individuals afflicted with  
C Alzheimer's disease. Inhibitors of beta-secretase are administered to a  
C mammalian subject e.g. with Alzheimer's disease or Alzheimer's disease-  
C like pathology to test if they maintain or improve cognitive ability or  
C reduce the plaque burden. The compounds are used for the treatment of  
C amyloidogenic diseases e.g. Alzheimer's disease. The present sequence  
C represents an inhibitor of beta-secretase enzyme

X Sequence 8 AA;

Query Match 100.0%; Score 19; DB 3; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
|  
b 5 VAEF 8

RESULT 4  
ABB56241

ID ABB56241 standard; peptide; 8 AA.

XX ABB56241;

AC ABB56241;

DT 15-FEB-2002 (first entry)

XX Vascular dementia-associated protein isoform (VPI) 441.

XX Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;  
KW diagnosis; prognosis; gene therapy.

OS Homo sapiens.

XX WO200169261-A2.

XX 20-SEP-2001.

PF 14-MAR-2001; 2001WO-GB001106.

XX 15-MAR-2000; 2000GB-00006285.

PR 24-NOV-2000; 2000GB-00028734.

XX 28-NOV-2000; 2000US-00724391.

PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

XX Herath HMAC, Parekh RB, Rohlf C;

XX WPI; 2001-557937/62.

XX Screening, diagnosis or prognosis of vascular dementia (VD), useful for

PT determining stage of VD and monitoring the effect of VD therapy;

PT comprises analyzing body fluid by 2-dimensional electrophoresis for

XX features correlated with VD.

XX Claim 6; Page 39; 151pp; English.

CC The invention relates to screening, diagnosis or prognosis of Vascular  
CC Dementia (VD) in a subject comprising analysing body fluid from the  
CC subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of  
CC features containing at least one chosen feature whose relative abundance  
CC correlates with the presence, absence, stage or severity of VD or  
CC predicts the onset or course of VD, especially detecting in a sample of  
CC cerebrospinal fluid (CSF) from the subject one of 223 VD-associated  
CC protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the  
CC specification. Detecting VD-associated features and VPI is useful for the  
CC screening, diagnosis or prognosis of VD, for determining the stage or  
CC severity of VD, for identifying a subject at risk of VD or for monitoring  
CC the effect of therapy administered to a subject having VD. Nucleic acids  
CC encoding a VPI or inhibiting the function of a VPI are useful for the  
CC treatment of VD and for gene therapy

SQ Sequence 8 AA;

Query Match 100.0%; Score 19; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|  
Db 2 VAEF 5

RESULT 5

AAU28720  
ID AAU28720 standard; peptide; 8 AA.

XX AAU28720;

XX 03-JAN-2002 (first entry)

DT

XX DE DPI tryptic digest peptide #317.  
 XX DE Human; depression associated protein isoform: tryptic digest peptide;  
 KW DP; cerebrospinal fluid; CSF; BAD; bipolar affective disorder;  
 KW neuropsychiatric disorder; bipolar mood disorder; neuroleptic;  
 KW manic-depressive illness; schizoaffective disorder.  
 XX OS Homo sapiens.  
 XX PN W0200162787-A1.  
 XX PD 30-AUG-2001.  
 XX PF 23-FEB-2001; 2001WO-GB000786.  
 XX PR 24-FEB-2000; 2000GB-00004412.  
 XX PR 08-DEC-2000; 2000GB-00030050.  
 XX PR 12-DEC-2000; 2000US-0254830P.  
 XX FA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
 XX PI Herath HMAC, Parekh RB, Rohlf C, Terrett JA, Tyson KL;  
 XX WIPI; 2001-570626/64.  
 XX PT Novel nucleic acid encoding a protein associated with bipolar affective  
 PT disorder, which is used for diagnosis, prophylaxis and therapy of  
 PT neuropsychiatric disorders, such as bipolar affective disorder.  
 XX PS Disclosure; Page 37; 153pp; English.  
 XX CC The present invention relates to the identification of depression  
 CC associated protein isoforms (DPIs), particularly the tryptic digest  
 CC peptides of these proteins. Some of the DPIs (AAU28404-AAU28625)  
 CC described are decreased in the cerebrospinal fluid (CSF) of BAD (bipolar  
 CC affective disorder) subjects, whilst other DPIs (AAU28626-AAU28887) are  
 CC increased in BAD subjects. Also described are peptide sequences  
 CC identified from DPI-45 and DPI-213 and the nucleic acid sequence they are  
 CC encoded by. The sequences of the invention are useful for clinical  
 CC screening, diagnosis, prognosis, therapy and prophylaxis of  
 CC neuropsychiatric disorders e.g. BAD (also known as bipolar mood disorder,  
 CC BP), manic-depressive illnesses, attention deficit disorders,  
 CC schizoaffective disorders, and unipolar affective disorders. The present  
 CC sequence represents one of the DPI tryptic digest peptides of the present  
 CC invention  
 XX CC  
 XX SQ Sequence 8 AA;  
 Query Match 100.0%; Score 19; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 |||||  
 Db 2 VAEF 5  
 RESULT 6  
 AAU25114  
 ID AAU25114 standard; peptide; 8 AA.  
 AC AAU25114;  
 XX DT 18-DEC-2001 (first entry)  
 XX DE Schizophrenia-Associated Protein Isoform (SPI) peptide #343.  
 XX KW Schizophrenia-associated protein isoform; SPI; SPI-206; SPI-238; SPI-240;  
 KW neuroleptic; gene therapy; cerebrospinal fluid; serum; plasma.  
 XX OS Homo sapiens.

PN W0200162785-A2.  
 XX PD 30-AUG-2001.  
 XX PF 23-FEB-2001; 2001WO-GB000792.  
 XX PR 24-FEB-2000; 2000GB-00004415.  
 XX PR 28-DEC-2000; 2000US-00750395.  
 XX FA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
 XX PI Herath HMAC, Parekh RB, Rohlf C, Terrett JA, Tyson KL;  
 XX WIPI; 2001-570624/64.  
 XX PT New schizophrenia associated protein isoforms and encoding nucleic acid  
 PT molecules, useful for treatment, diagnosis and prognosis of schizophrenia  
 PT and screening for potential drugs for treatment and new drug targets.  
 XX PS Disclosure; Page 36; 148pp; English.  
 XX CC The sequence represents a schizophrenia-associated protein isoform (SPI).  
 CC These protein isoforms, e.g. SPI-206, SPI-238 and SPI-240 are detectable  
 CC in cerebrospinal fluid, serum or plasma and are useful markers of  
 CC schizophrenia. The sequences can be used for treatment and diagnosis of  
 CC schizophrenia, screening, prognosis, monitoring the results of therapy,  
 CC identifying patients most likely to respond to a particular therapy and  
 CC identification of new targets for drug treatment. SPI DNA is useful as a  
 CC nucleic acid probe to detect the presence of nucleic acids or SPIs  
 XX CC  
 XX SQ Sequence 8 AA;  
 Query Match 100.0%; Score 19; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 |||||  
 Db 2 VAEF 5  
 RESULT 7  
 AAU26368  
 ID AAU26368 standard; peptide; 8 AA.  
 XX AC AAU26368;  
 XX DT 18-DEC-2001 (first entry)  
 XX DE Depression-Associated Protein Isoform DPI-124.  
 XX KW Human; Bipolar Affective Disorder; BAD; Depression-Associated feature;  
 KW DP; Depression-Associated protein isoform; DPI; Cerebro-spinal fluid;  
 KW CSF; antidepressant, antimanic; isotropic; tranquiliser; neuroleptic;  
 KW attention deficient disorder; schizoaffective disorder;  
 KW unipolar affective disorder.  
 XX OS Homo sapiens.  
 XX PN W0200163294-A2.  
 XX PD 30-AUG-2001.  
 XX PF 23-FEB-2001; 2001WO-GB000791.  
 XX PR 24-FEB-2000; 2000GB-00004412.  
 XX PR 08-DEC-2000; 2000GB-00030050.  
 XX PR 12-DEC-2000; 2000US-0254830P.  
 XX FA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
 XX PI Herath HMAC, Parekh RB, Rohlf C;  
 XX WIPI; 2001-570624/64.

R WPI; 2001-582081/65.  
 X Preparation for diagnosing or treating bipolar affected disorder (BAD) or  
 T unipolar depression, or for screening for modulators, comprises a BAD-  
 T associated protein isoform.  
 X  
 S Claim 8; Page 37; 163pp; English.  
 X  
 C The invention relates to a preparation comprising an isolated Bipolar  
 C Affected Disorder (BAD)-Associated Protein Isoform (DPIs). The DPI's are  
 C used to screen, diagnose or prognose of BAD or unipolar depression,  
 C determine the stage or severity of BAD or unipolar depression, identify a  
 C subject at risk of developing BAD or unipolar depression, or monitor the  
 C effect of therapy in a subject. They are also used to screen for or  
 C identify agents that interact with a DPI. These agents, antibodies  
 C against the DPIs, and nucleic acids encoding the DPIs are used to treat  
 C or prevent BAD or unipolar depression. Diseases that can be treated are  
 C attention deficient disorder, a schizoaffective disorder, a bipolar or a  
 C unipolar affective disorder. The DPIs are used in proteomics. The  
 C proteomic approach of using DPIs for screening, diagnosis or prognosis of  
 C BAD or unipolar depression overcomes the problems of using gene  
 C expression analysis, such as not being able to obtain central nervous  
 C system (CNS) tissue from a living patient under normal circumstances. The  
 C present sequence is a DIP increased in the CSF (cerebro-spinal fluid) of  
 C subjects having BAD  
 X  
 Q Sequence 8 AA;  
 Query Match 100.0%; Score 19; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 1 VAEF 4  
 b | | | |  
 2 VAEF 5  
 RESULT 8  
 AUI5458  
 D AAUI5458 standard; peptide; 8 AA.  
 X  
 C AAUI5458;  
 X  
 T 24-OCT-2001 (first entry)  
 X  
 E Schizophrenia-associated isoform peptide #343.  
 X  
 W Schizophrenia; neuroleptic; diagnostic; neuropsychiatric disorder;  
 W neurological disorder; neuropathy.  
 X  
 S Homo sapiens.  
 X  
 N WO200163293-A2.  
 X  
 D 30-AUG-2001.  
 X  
 F 23-FEB-2001; 2001WO-GS000783.  
 X  
 R 24-FEB-2000; 2000GB-00004415.  
 R  
 R 28-DEC-2000; 2000US-00750395.  
 X  
 A (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
 X  
 X Herath HWAC, Parekh RB, Rohlf C;  
 I  
 R WPI; 2001-502868/55.  
 X  
 X Diagnosing and monitoring Schizophrenia by detecting the presence of  
 T Schizophrenia Associated Features and Schizophrenia Associated Protein  
 T isoforms in samples of cerebrospinal fluid.  
 X  
 S Claim 6; Page 36; 160pp; English.  
 X

CC The invention relates to methods and compositions for screening,  
 CC diagnosis and prognosis of Schizophrenia. The method involves detecting  
 CC the presence of Schizophrenia (SCH) Associated Features (SfS) and SCH  
 CC Associated Protein Isoforms (SPIs) in samples, e.g. by electrophoresis,  
 CC immunoassay or hybridisation assay, for diagnosing and monitoring SCH,  
 CC studying the effectiveness of treatments and for identifying potential  
 CC therapeutic agents. The method is used for (1) screening or diagnosis of  
 CC SCH and the relative abundance of at least 1 chosen feature correlates  
 CC with the presence or absence of SCH; and (2) monitoring the effect of  
 CC therapy administered to a subject with SCH and the relative abundance of  
 CC at least 1 chosen feature which correlates with the severity of SCH. The  
 CC expression and activity of the SfS, SPIs and related molecules (e.g.  
 CC secondary messengers) are studied to diagnose SCH, monitor the progress  
 CC of the disorder and the effectiveness of treatment, and as targets to  
 CC identify and produce potential therapeutic agents for the treatment of  
 CC SCH. The paucity of detectable neurologic defects distinguishes  
 CC neuropsychiatric disorders such as SCH from neurological disorders, where  
 CC manifestations of anatomical and biochemical changes have been identified  
 CC in many cases. Consequently the identification and characterisation of  
 CC cellular and/or molecular causative defects and neuropathies are  
 CC necessary for improved treatment of neuropsychiatric disorders. AAUI5114-  
 CC AAUI5762 represent the amino acid sequences of schizophrenia-associated  
 CC isoforms used in the method of the invention  
 XX  
 SQ Sequence 8 AA;  
 Query Match 100.0%; Score 19; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 Db | | | |  
 2 VAEF 5  
 RESULT 9  
 AAR78909  
 ID AAR78909 standard; peptide; 9 AA.  
 X  
 AC AAR78909;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 27-MAR-1996 (first entry)  
 XX  
 DE MAGE 3 105-113 cytotoxic T lymphocyte epitope.  
 XX  
 KW MAGE 3 105-113; cytotoxic T; CTL; epitope; helper T; HTL; cell;  
 KW lymphocyte; antigens; treatment; disease prevention; tumours; cancer;  
 KW melanomas.  
 XX  
 OS Homo sapiens.  
 XX  
 PV WO9522317-A1.  
 XX  
 PD 24-AUG-1995.  
 XX  
 PF 16-FEB-1995; 95WO-US002121.  
 XX  
 PR 16-FEB-1994; 94US-00197484.  
 XX  
 PA (CYTE-) CYTEL CORP.  
 XX  
 PI Vitiello MA, Chesnut RW, Sette AD, Celis E, Grey H;  
 XX  
 DR WPI; 1995-302545/39.  
 XX  
 PT Compn. inducing cytotoxic T lymphocyte response to pref. viral,  
 PT bacterial, parasitic or tumour antigens - useful in the treatment and  
 PT prevention of diseases associated with the antigen e.g. hepatitis B.  
 XX  
 PS Example 13; Page 71; 109pp; English.  
 XX  
 CC A compsn. which induces a cytotoxic T lymphocyte (CTL) response to a

CC human MAGE antigen (Ag) in a mammal comprises, a MAGE CTL Ag response  
 CC inducing peptide (i.e. AAR78904 to AAR78917) and a lipid conjugated  
 CC helper T cell inducing peptide. The compsn. is useful in the treatment  
 CC and prevention of MAGE tumour Ag associated diseases, e.g. melanoma  
 CC cancers. (Updated on 25-MAR-2003 to correct PI field.)  
 XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 19; DB 2; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB ||||  
 2 VAEF 5

RESULT 10

ABB77871  
 ID ABB77871 standard; peptide; 9 AA.

XX AC ABB77871;  
 XX 27-SEP-2002 (first entry)  
 XX A beta-amyloid precursor protein (APP) inhibitor peptide.  
 DE  
 XX Beta-amyloid precursor protein; APP; APP inhibitor peptide; BACE;  
 XX beta site APP cleaving enzyme; protein coordinate data; APP751;  
 KW Swedish family mutation; Alzheimer's disease.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.

XX Key Location/Qualifiers  
 FH Misc-difference 5 /label= Sta  
 FT /note= "statine"  
 FT

PN WO200225276-A1.

XX 28-MAR-2002.

XX 19-SEP-2001; 2001WO-US029387.

XX 22-SEP-2000; 2000US-0234576P.

PA (AMHP ) AMERICAN HOME PROD CORP.

XX Choppa R, Svenson K, Annis B, Akopian TN, Bard J, Stahl ML;  
 PI Somers WS;

XX WPI; 2002-519081/55.

XX Crystallized complex of beta-site amyloid precursor protein (APP)  
 PT cleaving enzyme (BACE) and APP inhibitor peptide useful for identifying  
 PT agents that interact with active site of BACE or active site of APP  
 PT binding protein or peptide.

XX Claim 1; Page 22; 88pp; English.

XX The present sequence represents a beta-amyloid precursor protein (APP)  
 CC inhibitor peptide. This peptide inhibits binding between APP and BACE  
 CC (beta site APP cleaving enzyme). The specification describes a  
 CC crystallized complex of BACE and the present APP inhibitor. Protein  
 CC coordinate data for BACE is given in the specification. The APP inhibitor  
 CC peptide is based on the P10 to P4' APP751 Swedish family mutation. The  
 CC crystallized complex is used for identifying an agent that interacts with  
 CC an active site of BACE or an active site of an APP binding protein or  
 CC peptide. The agents are useful in the treatment and/or prevention of  
 CC Alzheimer's disease  
 XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 19; DB 5; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB ||||  
 6 VAEF 9

RESULT 11

ABB06593  
 ID ABB06593 standard; peptide; 10 AA.

XX AC ABB06593;

XX 31-MAY-2002 (first entry)

XX Beta-secretase related peptide SEQ ID NO:197.

XX Beta-secretase; enzyme; cleavage site; amyloid protein precursor; APP;  
 KW aspartyl protease; neuroprotective; nootropic; beta-secretase inhibitor;  
 KW Alzheimer's disease.

XX Homo sapiens.  
 OS Synthetic.

XX WO200206306-A2.

XX 24-JAN-2002.

XX 19-JUL-2001; 2001WO-US023035.

XX 19-JUL-2000; 2000US-0219795P.

XX 12-MAR-2001; 2001US-0275251P.

XX (PHAA ) PHARMACIA & UPJOHN CO.

XX Yan R, Tomaselli AG, Gurney ME, Emmons TL, Bienkowski MJ;  
 PI Heinrichson RL;

XX WPI; 2002-216995/27.

XX Novel substrates for human aspartyl protease useful for identifying  
 PT modulators of beta secretase activity of aspartyl protease for treating  
 PT Alzheimer's disease.

XX Disclosure; Page 188; 188pp; English.

XX The present invention describes an isolated peptide (I) comprising a  
 CC sequence of at least four amino acids, where the peptide is a substrate  
 CC for conducting aspartyl protease assays. (I) has neuroprotective and  
 CC nootropic activities, and can be used as an inhibitor of beta-secretase  
 CC activity. A beta-secretase modulator from the present invention can be  
 CC used for inhibiting beta-secretase activity in vivo and in the  
 CC manufacture of a medicament for the treatment of Alzheimer's disease.  
 CC Pharmaceutical compositions from the present invention can be used for  
 CC treating a disease or condition characterised by an abnormal beta-  
 CC secretase activity. (I) is useful for identifying agents that modulate  
 CC the activity of human Asp2 aspartyl protease (Hu-Asp2). (I) is useful as  
 CC a core structure to construct derivatives. ABL49914 to ABL49925 and  
 CC ABB06409 to ABB06593 represent sequences used in the exemplification of  
 CC the present invention  
 XX

SQ Sequence 10 AA;

Query Match 100.0%; Score 19; DB 5; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB ||||  
 5 VAEF 8

RESULT 12  
BB06592  
D ABB06592 standard; peptide; 12 AA.  
C ABB06592;  
X T 31-MAY-2002 (first entry)  
X X Beta-secretase related peptide SEQ ID NO:196.  
E Beta-secretase; enzyme; cleavage site; amyloid protein precursor; APP;  
W aspartyl protease; neuroprotective; nontropic; beta-secretase inhibitor;  
W Alzheimer's disease.  
X X Homo sapiens.  
S Synthetic.  
X X WO200206306-A2.  
N 24-JAN-2002.  
D 19-JUL-2001; 2001WO-US023035.  
F 19-JUL-2000; 2000US-0219795P.  
R 12-MAR-2001; 2001US-0275251P.  
X (PHAA ) PHARMACIA & UPJOHN CO.  
A Yan R, Tomasselli AG, Gurney ME, Emmons TL, Bienkowski MJ;  
I Heinrikson RL;  
I WPI; 2002-216995/27.  
R Novel substrates for human aspartyl protease useful for identifying  
I modulators of beta secretase activity of aspartyl protease for treating  
I Alzheimer's disease.  
X Disclosure; Page 188; 189pp; English.  
X The present invention describes an isolated peptide (I) comprising a  
C sequence of at least four amino acids, where the peptide is a substrate  
C for conducting aspartyl protease assays. (I) has neuroprotective and  
C nontropic activities, and can be used as an inhibitor of beta-secretase  
C activity. A beta-secretase modulator from the present invention can be  
C used for inhibiting beta-secretase activity in vivo, and in the  
C manufacture of a medicament for the treatment of Alzheimer's disease.  
C Pharmaceutical compositions from the present invention can be used for  
C treating a disease or condition characterised by an abnormal beta-  
C secretase activity. (I) is useful for identifying agents that modulate  
C the activity of human Asp2 aspartyl protease (Hu-Asp2). (I) is useful as  
C a core structure to construct derivatives. ABL49914 to ABL49925 and  
C ABB06409 to ABB06593 represent sequences used in the exemplification of  
C the present invention

Sequence 12 AA;  
Query Match 100.0%; Score 19; DB 5; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
b 5 VAEF 8

RESULT 13  
BG78404  
D ABG78404 standard; peptide; 12 AA.  
X  
C ABG78404;  
X

DT 15-NOV-2002 (first entry)  
XX Memapsin 2 substrate specificity determination peptide #10.  
DE Human; memapsin 2; aspartic protease; beta secretase;  
XX degenerative disease; Alzheimer's disease; amyloid precursor protein;  
KW APP; neuroprotective; nontropic; inhibitor;  
KW substrate side-chain preference.  
XX Homo sapiens.  
OS Synthetic.  
XX WO200253594-A2.  
PN 11-JUL-2002.  
XX 28-DEC-2001; 2001WO-US050826.  
PF 28-DEC-2000; 2000US-0258705P.  
PR 14-MAR-2001; 2001US-0275756P.  
XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
PA (UNII ) UNIV ILLINOIS FOUND.  
PA Tang JUN, Koelsch G, Ghosh AK;  
PI WPI; 2002-619089/66.  
XX New memapsin 2 activity inhibitor useful in treatment of e.g. Alzheimer's  
PT disease.  
PT Example 5; Page 51; 74pp; English.

XX The invention relates to an inhibitor of catalytically active memapsin 2  
CC (an aspartic protease which can cleave at beta secretase sites), which  
CC binds to the active site of memapsin 2 defined by the presence of two  
CC catalytic aspartic residues and substrate binding cleft. Also included is  
CC a method of determination of the substrate side-chain preference in  
CC memapsin 2 sub-sites comprising: (a) reacting a mixture of memapsin 2  
CC substrates with memapsin 2, and determining the sub-site preference of  
CC memapsin 2 by determining relative initial hydrolysis rates of the  
CC mixture of memapsin 2 substrates; or (b) preparing a combinatorial  
CC library of memapsin 2 inhibitors containing a base sequence taken from  
CC OM99-2 (Glu-Val-Asn-Leu-Ala-Glu-phe), probing the library of  
CC inhibitors with memapsin 2 which binds to several inhibitors to generate  
CC several bound memapsin 2, and detecting the bound memapsin 2 with an  
CC antibody raised to memapsin 2, and an alkaline phosphatase conjugated  
CC secondary antibody. The inhibitors may be used in the manufacture of a  
CC medicament for the treatment of Alzheimer's disease since memapsin 2 may  
CC be involved in the cleavage of amyloid precursor protein (APP), and for  
CC determining the substrate side-chain preference in memapsin 2 sub-sites.  
CC The present sequence represents a sub-site variant peptide used to  
CC determine the substrate specificity of human memapsin 2

Sequence 12 AA;

Query Match 100.0%; Score 19; DB 5; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 5 VAEF 8

RESULT 14  
ADD35467  
ID ADD35467 standard; peptide; 12 AA.  
XX  
AC ADD35467;  
XX  
DT 15-JAN-2004 (first entry)  
XX

DE Escherichia coli DnaX peptide #3.  
 XX  
 XX crystallised recombinant protein; metabolism; Staphylococcus aureus;  
 KW Streptococcus pneumoniae; Helicobacter pylori; Escherichia coli;  
 KW Pseudomonas aeruginosa; vaccine.  
 XX  
 OS Escherichia coli.  
 XX  
 XX WO2003044185-A2.  
 PN  
 XX  
 XX 30-MAY-2003.  
 PD  
 XX  
 XX 21-NOV-2002; 2002WO-CA001768.  
 PF  
 XX  
 XX 21-NOV-2001; 2001US-0332160P.  
 PR  
 XX 27-NOV-2001; 2001US-0333661P.  
 PR  
 XX 27-NOV-2001; 2001US-0333665P.  
 PR  
 XX 18-DEC-2001; 2001US-0341770P.  
 PR  
 XX 19-DEC-2001; 2001US-0341954P.  
 PR  
 XX 19-DEC-2001; 2001US-0342003P.  
 PR  
 XX 20-DEC-2001; 2001US-0342542P.  
 PR  
 XX 21-DEC-2001; 2001US-0344252P.  
 PR  
 XX 28-DEC-2001; 2001US-0343570P.  
 PR  
 XX 28-DEC-2001; 2001US-0343606P.  
 PR  
 XX 28-DEC-2001; 2001US-0343679P.  
 PR  
 XX  
 PA (AFFI-) AFFINIUM PHARM INC.  
 XX  
 XX Edwards A, Dharamsi A, Vedadi M, Alam MZ, Awrey D, Beattie B;  
 PI Canadian V, Domagala M, Houston S, Mansoury K, Necakov S, Nethery K;  
 PI Ng I, Pinder B, Sheldrick B, Vallee F, Wrezel O;  
 XX  
 XX WPI; 2003-513596/48.  
 DR  
 XX New crystallized recombinant polypeptides from Staphylococcus aureus,  
 PT Streptococcus pneumoniae, Helicobacter pylori or Pseudomonas aeruginosa  
 PT involved in general metabolism, useful as drug targets for pathogenic  
 PT bacteria.  
 XX  
 PS Disclosure; SEQ ID NO 66; 277pp; English.  
 XX  
 CC The invention comprises a crystallised recombinant protein that is  
 CC involved in general metabolism, the recombinant protein may be from  
 CC Staphylococcus aureus, Streptococcus pneumoniae, Helicobacter pylori,  
 CC Escherichia coli or Pseudomonas aeruginosa. The crystallised recombinant  
 CC protein of the invention is useful in the prevention (vaccine) or  
 CC treatment of a disease or disorder caused by S. pneumoniae, H. pylori, E.  
 CC coli or P. aeruginosa. The present amino acid sequence was used in the  
 CC exemplification of the invention.  
 XX  
 SQ Sequence 12 AA;  
 XX  
 Query Match 100.0%; Score 19; DB 7; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 ||||  
 Db 9 VAEF 12  
 RESULT 15  
 AAB07889  
 ID AAB07889 standard; peptide; 13 AA.  
 XX  
 AC AAB07889;  
 XX  
 XX 14-NOV-2000 (first entry)  
 DT  
 XX  
 DE A beta-secretase inhibitor peptide.  
 XX  
 XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;

KW inhibitor.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200047618-A2.  
 XX  
 XX 17-AUG-2000.  
 PD  
 XX  
 XX 10-FEB-2000; 2000WO-US003819.  
 PF  
 XX  
 XX 10-FEB-1999; 99US-0119571P.  
 PR  
 XX 15-JUN-1999; 99US-0139172P.  
 PR  
 XX (ELAN-) ELAN PHARM INC.  
 PA  
 XX Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 PI Sinha S, Tatsuno G, Tung J, Wang S, Mcconlogue L;  
 PI  
 XX WPI; 2000-533011/48.  
 DR  
 XX Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease.  
 XX  
 XX Claim 33; Page 24; 121pp; English.  
 PS  
 XX  
 CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to a  
 CC mammalian subject e.g. with Alzheimer's disease or Alzheimer's disease-  
 CC like pathology to test if they maintain or improve cognitive ability or  
 CC reduce the plaque burden. The compounds are used for the treatment of  
 CC amyloidogenic diseases e.g. Alzheimer's disease. The present sequence  
 CC represents an inhibitor of beta-secretase enzyme  
 XX  
 SQ Sequence 13 AA;  
 XX  
 Query Match 100.0%; Score 19; DB 3; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 ||||  
 Db 10 VAEF 13  
 RESULT 16  
 AAM99276  
 ID AAM99276 standard; peptide; 13 AA.  
 XX  
 AC AAM99276;  
 XX  
 XX 07-DEC-2001 (first entry)  
 DT  
 XX  
 DE Vaccine related MHC ligand peptide SEQ ID NO:379.  
 XX  
 XX Glutamic acid; glutamine; vaccine; major histocompatibility complex; MHC;  
 KW immunomodulator; antiallergic; endocrine; neuroprotectant; virucidal;  
 KW bactericidal; antiparasitic; fungicidal; cytostatic; medicine;  
 KW pharmaceutical; immune disorder; immune deficiency; autoimmune;  
 KW hypersensitivity; allergy; graft rejection; infection; hormonal disorder;  
 KW central nervous system disease; cancer; melanoma; anti-melanoma vaccine;  
 KW human immunodeficiency virus.  
 XX  
 OS Mycobacterium leprae.  
 XX  
 XX WO200170772-A2.  
 PN  
 XX  
 XX 27-SEP-2001.  
 PD  
 XX  
 XX 22-MAR-2001; 2001WO-FR000872.

XX 23-MAR-2000; 2000FR-00003711.  
 XX (FABR ) FABRE MEDICAMENT SA PIERRE.  
 XX Klinguer-Hamour C, Corvaia N, Beck A, Goetsch L;  
 XX WPI; 2001-611470/70.  
 XX Stabilized pharmaceutical containing N-terminal glutamic acid or  
 XX glutamine, useful e.g. in anti-melanoma vaccines, is an addition salt  
 XX with strong acid.  
 XX Claim 9; Page 96; 149pp; French.  
 XX The present invention describes a pharmaceutical compound (I) that  
 XX contains an N-terminal glutamic acid (Glu) or glutamine (Gln) residue in  
 XX the form of an addition salt with a strong, physiologically acceptable  
 XX acid (II). Also described are: (a) a pharmaceutical composition  
 XX containing at least one (I); (b) a vaccine containing at least one (I)  
 XX where this is a major histocompatibility complex (MHC) ligand (Ia); (c) a  
 XX method for in vitro diagnosis of diseases associated with the presence of  
 XX (Ia); (d) a kit for method (c) that includes a (Ia); and (e) a process  
 XX for preparing (I). (I) has immunomodulator, endocrine, antiallergic,  
 XX neuroprotectant, virucidal, bactericidal, antiparasitic, fungicidal and  
 XX cytostatic activities. (I) are useful, in human or veterinary medicine,  
 XX in pharmaceutical compositions (for treating immune disorders, e.g.  
 XX immune deficiency, autoimmune states, hypersensitivity, allergy, graft  
 XX rejection, infection, hormonal disorders and central nervous system  
 XX diseases), also, where (I) is a MHC ligand (Ia), in vaccines for  
 XX treatment or prevention of: (i) viral, bacterial, parasitic or fungal  
 XX infections; or (ii) of cancers. A particular application is in anti-  
 XX melanoma vaccines. (I) are also useful for in vitro diagnosis of diseases  
 XX associated with interactions between MHC and (I), e.g. melanoma and human  
 XX immunodeficiency virus infection. AAM9898 to AAM99592 represent peptides  
 XX which can be used in pharmaceutical compounds from the present invention  
 XX  
 XX Sequence 13 AA;  
 XX  
 XX Query Match 100.0%; Score 19; DB 4; Length 13;  
 XX Best Local Similarity 100.0%; Pred. NO. 1.6e+02;  
 XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX 1 VAEF 4  
 XX ||||  
 XX 4 VAEF 7  
 XX  
 XX RESULT 17  
 XX LAB07888  
 XX D AAB07888 standard; peptide; 14 AA.  
 XX  
 XX AAB07888;  
 XX  
 XX 14-NOV-2000 (first entry)  
 XX  
 XX A peptide fragment derived from beta-amyloid precursor protein.  
 XX  
 XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 XX amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 XX inhibitor.  
 XX Homo sapiens.  
 XX WO2000047618-A2.  
 XX  
 XX 17-AUG-2000.  
 XX  
 XX 10-FEB-2000; 2000WO-US003819.  
 XX  
 XX 10-FEB-1999; 99US-0119571P.  
 XX 15-JUN-1999; 99US-0139172P.  
 XX

PA (ELAN-) ELAN PHARM INC.  
 XX Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 XX Sinha S, Tatsuno G, Tung J, Wang S, Mcconlogue L;  
 XX WPI; 2000-533011/48.  
 XX Purified beta-secretase protein used in assays to discover inhibitors  
 XX which can be used for the treatment of amyloidogenic diseases e.g.  
 XX Alzheimer's disease.  
 XX Disclosure; Page 12; 121pp; English.  
 XX  
 XX The specification describes a beta-secretase enzyme. The enzyme cleaves  
 XX beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 XX enzyme is therefore implicated in the production of amyloid plaque  
 XX components which accumulate in the brains of individuals afflicted with  
 XX Alzheimer's disease. Inhibitors of beta-secretase are administered to a  
 XX mammalian subject e.g. with Alzheimer's disease or Alzheimer's disease-  
 XX like pathology to test if they maintain or improve cognitive ability or  
 XX reduce the plaque burden. The compounds are used for the treatment of  
 XX amyloidogenic diseases e.g. Alzheimer's disease. The present sequence  
 XX represents a peptide derived from beta-amyloid precursor protein  
 XX  
 XX Sequence 14 AA;  
 XX  
 XX Query Match 100.0%; Score 19; DB 3; Length 14;  
 XX Best Local Similarity 100.0%; Pred. NO. 1.8e+02;  
 XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX 1 VAEF 4  
 XX ||||  
 XX 11 VAEF 14  
 XX  
 XX RESULT 18  
 XX AAM97460  
 XX ID AAM97460 standard; peptide; 14 AA.  
 XX  
 XX AAM97460;  
 XX  
 XX 24-JAN-2002 (first entry)  
 XX  
 XX Human peptide #735 encoded by a SNP oligonucleotide.  
 XX  
 XX Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;  
 XX neuroprotective; antimicrobial; gene therapy; vaccine; amyase; cancer;  
 XX amyloid protein; angiotensin; apoptosis related protein; cadherin;  
 XX cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;  
 XX complement related protein; cytochrome; kinesin; cytokine; interferon;  
 XX interleukin; G-protein coupled receptor; thicesterase; inflammation;  
 XX multifactorial disease; autoimmune disease; infection;  
 XX nervous system disease.  
 XX Homo sapiens.  
 XX WO200147944-A2.  
 XX  
 XX 05-JUL-2001.  
 XX  
 XX 28-DEC-2000; 2000WO-US035498.  
 XX  
 XX 28-DEC-1999; 99US-0173419P.  
 XX 27-DEC-2000; 2000US-00173419.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Shimkets RA, Leach M;  
 XX WPI; 2001-465210/50.  
 XX  
 XX Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
 XX oncogenes and histones, useful for diagnosing and treating, e.g. cancer,  
 XX

PT autoimmune diseases and infections.  
 XX Disclosure; Page 3829; 4143pp; English.  
 PS The present invention relates to oligonucleotides (see AAL26793-AAL34659)  
 XX encoding polymorphic variants of proteins related to amylases, amyloid  
 CC proteins, angiotensin, apoptosis related proteins, cadherin, cyclin,  
 CC polymerase, oncogenes, histones, kinases, colony stimulating factors,  
 CC complement related proteins, cytochromes, kinesins, cytokines,  
 CC interferons, interleukins, G-protein coupled receptors and thioesterases.  
 CC The present sequence is a peptide encoded by one such oligonucleotide.  
 CC The oligonucleotides and the peptides encoded by them may be used in the  
 CC prevention, diagnosis and treatment of diseases associated with  
 CC inappropriate expression of the proteins listed above. Disorders that may  
 CC be prevented, diagnosed and/or treated include multifactorial diseases  
 CC with a genetic component, such as autoimmune diseases (e.g. rheumatoid  
 CC arthritis, multiple sclerosis, diabetes, systemic lupus erythematosus  
 CC and Grave's disease), inflammation, cancer (e.g. cancers of the bladder,  
 CC brain, breast, colon and kidney, leukaemia), diseases of the nervous  
 CC system and an infection of pathogenic organisms  
 XX Sequence 14 AA;  
 SQ

Query Match 100.0%; Score 19; DB 4; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 |||||  
 DB 5 VAEF 8

RESULT 19  
 AAE32223  
 ID AAE32223 standard; peptide; 15 AA.  
 XX  
 AC AAE32223;  
 DT 24-MAR-2003 (first entry)  
 DE Arabidopsis thaliana S11 peptide #6.  
 XX Genomic database; mass spectrometer; proteomic business; pharmaceutical;  
 KW nuclear transport; signalling pathway; cellular organelle.  
 XX Arabidopsis thaliana.  
 OS WO200280649-A2.  
 PN 17-OCT-2002.  
 PD 09-APR-2002; 2002WO-US011417.  
 XX 09-APR-2001; 2001US-0282551P.  
 PR 20-APR-2001; 2001US-0285362P.  
 XX (MDS- ) MDS PROTEOMICS INC.  
 PA Mann M, Mortensen P;  
 XX WPI; 2003-058452/05.  
 DR Identifying coding sequence in genomic databases for conducting  
 PT proteomics business, by using polypeptide sequence information obtained  
 PT from peptide sequencing projects, especially those using mass  
 PT spectrometers.  
 PS Example 1; Page 41; 100pp; English.  
 XX The invention relates to a method for identifying coding sequence in  
 CC genomic databases, by using polypeptide sequence information obtained  
 CC from peptide sequencing projects, especially those using mass  
 CC spectrometers. It is useful for conducting a proteomics business. It is

CC also used for establishing a distribution system for distributing the  
 CC pharmaceutical preparation for sale, and may optionally include  
 CC establishing a sales group for marketing the pharmaceutical preparation.  
 CC It is also used for predicting the gene structure, such as intron/exon  
 CC boundaries, for searching genomic databases for sequences derived from  
 CC multi-protein complexes e.g. assemblies with a particular function such  
 CC as splicing, transport or nuclear import or export; for elucidating  
 CC transient rather than structural complexes (that are involved in  
 CC signalling pathways) and for identifying proteins in cellular organelles.  
 CC The present sequence is Arabidopsis thaliana peptide used to illustrate  
 CC the method of the invention  
 XX Sequence 15 AA;  
 SQ

Query Match 100.0%; Score 19; DB 6; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 |||||  
 DB 10 VAEF 13

RESULT 20  
 AAW42944  
 ID AAW42944 standard; peptide; 20 AA.  
 XX  
 AC AAW42944;  
 DT 28-APR-1998 (first entry)  
 XX Immunogenic Hepatitis A virus peptide YK-1369.  
 DE Immunogenic peptide; immunogenic epitope; P3A protein; immune response;  
 KW antibody.  
 XX Synthetic.  
 OS Hepatitis A virus.  
 XX WO9740147-A1.  
 PN 30-OCT-1997.  
 PD 18-APR-1997; 97WO-US006891.  
 XX 19-APR-1996; 96US-0015644P.  
 PR (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Fields HA, Khudyakov YE;  
 PI WPI; 1997-535831/49.  
 DR Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
 PT response to HAV in a mammal or to detect the presence of antibodies  
 PT against HAV in a mammal.  
 XX Claim 33; Page 115; 140pp; English.  
 PS Peptides AAW42943-46 are immunogenic peptides corresponding to  
 CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
 CC substantially similar to a portion of the amino acid sequence of the P3A  
 CC protein of HAV corresponding to amino acids 1423-1496. The present  
 CC peptide is derived from amino acids 1430-1449 and has a reactivity of  
 CC 58.5% with acute sera. Compositions containing the peptides can be used  
 CC to induce an immune response to HAV in a mammal. The peptides can also be  
 CC used to detect the presence of antibodies against HAV in mammalian serum.  
 CC The peptides can also be used to make an antibody against HAV by  
 CC administering the peptide to a mammal  
 XX Sequence 20 AA;  
 SQ

Query Match 100.0%; Score 19; DB 2; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.6e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
1  
1  
1  
1  
4 VAEF 7

RESULT 21  
AAW42943  
D AAW42943 standard; peptide; 20 AA.  
C AAW42943;  
X 28-APR-1998 (first entry)  
X Immunogenic Hepatitis A virus peptide YK-1368.  
X Immunogenic peptide; immunogenic epitope; P3A protein; immune response;  
X antibody.  
X Synthetic.  
X Hepatitis A virus.  
X W09740147-A1.  
X 30-OCT-1997.  
X 18-APR-1997; 97WO-US006891.  
X 19-APR-1996; 96US-0015644P.  
X (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
X Fields HA, Khudyakov YE;  
X WPI; 1997-535831/49.  
X Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
X response to HAV in a mammal or to detect the presence of antibodies  
X against HAV in a mammal.  
X Claim 33; Page 115; 140pp; English.  
X Peptides AAW42943-46 are immunogenic peptides corresponding to  
X immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
X substantially similar to a portion of the amino acid sequence of the P3A  
X protein of HAV corresponding to amino acids 1423-1496. The present  
X peptide is derived from amino acids 1421-1440 and has a reactivity of  
X 63.4% with acute sera. Compositions containing the peptides can be used  
X to induce an immune response to HAV in a mammal. The peptides can also be  
X used to detect the presence of antibodies against HAV in mammalian serum.  
X The peptides can also be used to make an antibody against HAV by  
X administering the peptide to a mammal

Sequence 20 AA;  
Query Match 100.0%; Score 19; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
1  
1  
1  
13 VAEF 16

RESULT 22  
AAB69462  
D AAB69462 standard; peptide; 21 AA.  
X AAB69462;  
X 20-APR-2001 (first entry)

Synthetic HAV P3A peptide, SEQ ID NO: 62.  
Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
antigen; major structural capsid polypeptide; HAV antibody detection.  
Hepatitis A virus.  
Synthetic.  
W0200105824-A2.  
25-JAN-2001.  
14-JUL-2000; 2000WO-US019267.  
15-JUL-1999; 99US-0144412P.  
(USSH ) US DEPT HEALTH & HUMAN SERVICES.  
Fields HA, Khudyakov YE;  
WPI; 2001-112681/12.  
Synthetic peptides used as antigen sources for enzyme immunoassays  
detecting anti-hepatitis A virus and as vaccines.  
Claim 22; Page 108; 130pp; English.  
The present sequence is one of a number of synthetic peptides which are  
immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
comprise antigenic epitopes of the major structural capsid polypeptides  
or non-structural polypeptides of HAV with one or more glutamine  
molecules at the carboxy end of the peptide. The peptides are used to  
detect the presence of antibodies against HAV in mammalian serum, to  
detect the presence of HAV in a human or animal through the binding of  
the peptide to an antibody, to detect acute phase infection by detecting  
IGM antibodies in mammalian serum and detecting convalescence in a  
mammal. The peptides are used to detect or quantify HAV antibodies in  
samples in clinical or research-based assays using immunoblotting,  
fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
tracking of radioactive or bioluminescent markers, chromatography or  
electrophoresis. The peptides are used to induce an immune response to  
HAV when administered to a human or animal. Glutamine at the carboxy end  
of the peptides enhances the IGM antibody reactivity

Sequence 21 AA;  
Query Match 100.0%; Score 19; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
1  
1  
1  
13 VAEF 16

RESULT 23  
AAB69463  
D AAB69463 standard; peptide; 21 AA.  
X AAB69463;  
X 20-APR-2001 (first entry)  
X Synthetic HAV P3A peptide, SEQ ID NO: 63.  
X Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
X antigen; major structural capsid polypeptide; HAV antibody detection.  
X Hepatitis A virus.  
X Synthetic.  
X W0200105824-A2.

XX PD 25-JAN-2001.  
 XX PF 14-JUL-2000; 2000WO-US019267.  
 XX PR 15-JUL-1999; 99US-0144412P.  
 XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX PI Fields HA, Khudyakov YE;  
 XX PI WPI; 2001-112681/12.  
 XX PT Synthetic peptides used as antigen sources for enzyme immunoassays  
 XX PT detecting anti-hepatitis A virus and as vaccines.  
 XX PS Claim 22; Page 109; 130pp; English.  
 XX CC The present sequence is one of a number of synthetic peptides which are  
 CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
 CC comprise antigenic epitopes of the major structural capsid polypeptides  
 CC or non-structural polypeptides of HAV with one or more glutamine  
 CC molecules at the carboxy end of the peptide. The peptides are used to  
 CC detect the presence of antibodies against HAV in mammalian serum, to  
 CC detect the presence of HAV in a human or animal through the binding of  
 CC the peptide to an antibody, to detect acute phase infection by detecting  
 CC Igm antibodies in mammalian serum and detecting convalescence in a  
 CC mammal. The peptides are used to detect or quantify HAV antibodies in  
 CC samples in clinical or research-based assays using immunoblotting,  
 CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
 CC tracking of radioactive or bioluminescent markers, chromatography or  
 CC electrophoresis. The peptides are used to induce an immune response to  
 CC HAV when administered to a human or animal. Glutamine at the carboxy end  
 CC of the peptides enhances the Igm antibody reactivity  
 XX SQ Sequence 21 AA;  
 XX Query Match 100.0%; Score 19; DB 4; Length 21;  
 XX Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX QY 1 VAEFF 4  
 XX Db 4 VAEFF 7  
 XX RESULT 24  
 XX AAW42946  
 XX ID AAW42946 standard; peptide; 25 AA.  
 XX AC AAW42946;  
 XX DT 28-APR-1998 (first entry)  
 XX DE Immunogenic Hepatitis A virus peptide YK-1832.  
 XX KW Immunogenic peptide; immunogenic epitope; P3A protein; immune response;  
 XX KW antibody.  
 XX OS Synthetic.  
 XX OS Hepatitis A virus.  
 XX PN WO9740147-A1.  
 XX PD 30-OCT-1997.  
 XX PF 18-APR-1997; 97WO-US006891.  
 XX PR 19-APR-1996; 96US-0015644P.  
 XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX PI Fields HA, Khudyakov YE;

XX WPI; 1997-535831/49.  
 XX PF Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
 XX PT response to HAV in a mammal or to detect the presence of antibodies  
 XX PT against HAV in a mammal.  
 XX PS Claim 33; Page 115; 140pp; English.  
 XX CC The present immunogenic peptide corresponds to an immunogenic epitope of  
 CC the Hepatitis A virus (HAV). The peptide is substantially similar to a  
 CC portion of the amino acid sequence of the P3A protein of HAV  
 CC corresponding to amino acids 1423-1496. Compositions containing the  
 CC peptide can be used to induce an immune response to HAV in a mammal. The  
 CC peptide can also be used to detect the presence of antibodies against HAV  
 CC in mammalian serum. The peptide can also be used to make an antibody  
 CC against HAV by administering the peptide to a mammal  
 XX SQ Sequence 25 AA;  
 XX Query Match 100.0%; Score 19; DB 2; Length 25;  
 XX Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX QY 1 VAEFF 4  
 XX Db 3 VAEFF 6  
 XX RESULT 25  
 XX AAB69465  
 XX ID AAB69465 standard; peptide; 25 AA.  
 XX AC AAB69465;  
 XX DT 20-APR-2001 (first entry)  
 XX DE Synthetic HAV P3A peptide, SEQ ID NO: 65.  
 XX KW Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
 XX KW antigen; major structural capsid polypeptide; HAV antibody detection.  
 XX OS Hepatitis A virus.  
 XX OS Synthetic.  
 XX PN WO200105824-A2.  
 XX PD 25-JAN-2001.  
 XX PF 14-JUL-2000; 2000WO-US019267.  
 XX PR 15-JUL-1999; 99US-0144412P.  
 XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX PI Fields HA, Khudyakov YE;  
 XX PI WPI; 2001-112681/12.  
 XX PT Synthetic peptides used as antigen sources for enzyme immunoassays  
 XX PT detecting anti-hepatitis A virus and as vaccines.  
 XX PS Claim 22; Page 110; 130pp; English.  
 XX CC The present sequence is one of a number of synthetic peptides which are  
 CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
 CC comprise antigenic epitopes of the major structural capsid polypeptides  
 CC or non-structural polypeptides of HAV with one or more glutamine  
 CC molecules at the carboxy end of the peptide. The peptides are used to  
 CC detect the presence of antibodies against HAV in mammalian serum, to  
 CC detect the presence of HAV in a human or animal through the binding of  
 CC the peptide to an antibody, to detect acute phase infection by detecting  
 CC Igm antibodies in mammalian serum and detecting convalescence in a

CC mammal. The peptides are used to detect or quantify HAV antibodies in  
 CC samples in clinical or research-based assays using immunoblotting,  
 CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
 CC tracking of radioactive or bioluminescent markers, chromatography or  
 CC electrophoresis. The peptides are used to induce an immune response to  
 CC HAV when administered to a human or animal. Glutamine at the carboxy end  
 CC of the peptides enhances the Igm antibody reactivity

XX Sequence 25 AA;

Query Match 100.0%; Score 19; DB 4; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
 ||||  
 3 VAEF 6

# RESULT 26

AAW17832 standard; peptide; 28 AA.

AAW17832;

29-JUL-1997 (first entry)

Human mevalonate pyrophosphate decarboxylase peptide NT71.

Mevalonate pyrophosphate decarboxylase; MPD; cholesterol.

Homo sapiens.

WO9714787-A1.

24-APR-1997.

10-OCT-1996; 96WO-EP004394.

18-OCT-1995; 95US-0005652P.

(NOVS) NOVARTIS AG.

Toth MJ, Huwyler LR;

WPI; 1997-245104/22.

Human mevalonate pyrophosphate decarboxylase coding sequence - used for  
 screening for MPD inhibitors, which regulate and control cholesterol  
 synthesis.

Example 3; Page 10; 37pp; English.

Peptide NT71 (AAW17832) was obtd. by trypsin digestion of human liver  
 mevalonate pyrophosphate decarboxylase (MPD) (AAW17831), an enzyme of the  
 cholesterol biosynthetic pathway. The sequence of the peptide was used to  
 design PCR primers utilised in the amplification of cDNA from a rat liver  
 cDNA library. A rat MPD partial clone was obtd. and used as a probe to  
 screen a human liver cDNA library. A 1800 bp sequence (AAW166464) coding  
 for human MPD was identified

Sequence 28 AA;

Query Match 100.0%; Score 19; DB 2; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
 ||||  
 20 VAEF 23

# RESULT 27

AAW33910 standard; protein; 28 AA.

AAW33910;

17-OCT-2001 (first entry)

Peptide #7947 encoded by probe for measuring placental gene expression.

Probe; microarray; human; placenta; antenatal diagnosis;

Genetic disorder.

Homo sapiens.

WO200157272-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US000663.

04-FEB-2000; 2000US-0180312P.

26-MAY-2000; 2000US-0207456P.

30-JUN-2000; 2000US-00608408.

03-AUG-2000; 2000US-00632366.

21-SEP-2000; 2000US-0234687P.

04-OCT-2000; 2000US-0236359P.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488897/53.

Human genome-derived single exon nucleic acid probes useful for analyzing  
 gene expression in human placenta.

Claim 27; SEQ ID NO 34179; 654pp; English.

The present invention relates to single exon nucleic acid probes (SENP:  
 see AA131315-AA157546). The present sequence is a peptide encoded by one  
 such probe. The probes are useful for producing a microarray for  
 predicting, measuring and displaying gene expression in samples derived  
 from human placenta. The probes are useful for antenatal diagnosis of  
 human genetic disorders

Sequence 28 AA;

Query Match 100.0%; Score 19; DB 4; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
 ||||  
 4 VAEF 7

# RESULT 28

ABG55464

ID ABG55464 standard; peptide; 28 AA.

ABG55464;

25-FEB-2003 (first entry)

Human liver peptide, SEQ ID No 34112.

Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;

hypercholesterolaemia; coronary heart disease.

Homo sapiens.

WO200157273-A2.



```

RESULT 30
D AAW62760 standard; protein; 30 AA.
X C AAW62760;
X T 09-NOV-1998 (first entry)
X E Streptococcus pneumoniae polypeptide.
X M Polypeptide; ORF; open reading frame; infection; bacterial;
X W streptococcal; bacteremia; diagnosis; prophylaxis.
X S Streptococcus pneumoniae.
X N WO9823631-A1.
X D 04-JUN-1998.
X F 24-NOV-1997; 97WO-US021976.
X R 27-NOV-1996; 96US-0031879P.
X A (SMIK ) SMITHKLINE BEECHAM CORP.
X A (SMIK ) SMITHKLINE BEECHAM PLC.
X X Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
X I Reid RH, Zarfos FN;
X R WPI; 1998-322654/28.
X S Streptococcus pneumoniae polynucleotides - useful for developing products
X C for diagnosis, prevention and treatment of infections e.g. pneumonia,
X C bacteremia, meningitis or endocarditis.
X S Claim 5; Page 32; 18pp; English.
X C The sequence is that of a Streptococcal polypeptide. The polypeptide can
X C potentially be used for the diagnosis and prevention of bacterial
X C infections, especially sp infection. It may be used for the treatment of
X C diseases such as otitis media, conjunctivitis, pneumonia, bacteremia,
X C meningitis, sinusitis, pleural empyema, endocarditis or infection of the
X C cerebrospinal fluid
X X Sequence 30 AA;
Query Match 100.0%; Score 19; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 4e+02; 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
ly 1 VAEF 4
lb ||||
21 VAEF 24
RESULT 31
D AAW82654 standard; protein; 32 AA.
X C AAW82654;
X T 07-NOV-2001 (first entry)
X E Human immune/haematopoietic antigen SEQ ID NO:10247.
X W Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
X C cytostatic; gene therapy; vaccine; metastasis.
X S Homo sapiens.
X X WO200157182-A2.
X X

```

```

PD 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US001354.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 28-JUN-2000; 2000US-0209467P.
PR 30-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0215647P.
PR 07-JUL-2000; 2000US-0218880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 05-SEP-2000; 2000US-0230437P.
PR 05-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234597P.
PR 25-SEP-2000; 2000US-0234598P.
PR 26-SEP-2000; 2000US-0234584P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.

```

PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241222P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246522P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0248526P.  
PR 08-NOV-2000; 2000US-0248527P.  
PR 08-NOV-2000; 2000US-0248528P.  
PR 08-NOV-2000; 2000US-0248532P.  
PR 08-NOV-2000; 2000US-0248533P.  
PR 08-NOV-2000; 2000US-0248609P.  
PR 08-NOV-2000; 2000US-0248610P.  
PR 08-NOV-2000; 2000US-0248611P.  
PR 08-NOV-2000; 2000US-0248613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249219P.  
PR 17-NOV-2000; 2000US-0249224P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 08-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
(HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-483426/52.  
XX

DR N-PSDB; AAK55435.  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
PS Claim 11; SEQ ID NO 10247; 3071pp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 32 AA;  
Query Match 100.0%; Score 19; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 4.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
Db 13 VAEF 16  
RESULT 32  
AAO09024  
ID AAO09024 standard; protein; 33 AA.  
XX  
AC AAO09024;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 22916.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US004927.  
XX  
PR 28-FEB-2000; 2000US-00515126.  
PR 18-MAY-2000; 2000US-00577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
PI WPI; 2001-514838/56.  
DR N-PSDB; AAI88955.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
PT and treating e.g. leukemia, inflammation and immune disorders.  
XX  
PS Claim 20; SEQ ID NO 22916; 1399pp + Sequence Listing; English.

X C The invention relates to human polynucleotides (AA179941-AA193841) and  
 C the encoded proteins (AA00010-AA013910) that exhibit activity relating to  
 C cytokine, cell proliferation or cell differentiation or which may induce  
 C production of other cytokines in other cell populations. The  
 C polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 C peptide therapy. The polypeptides have various cytokine-like activities,  
 C e.g. stem cell growth factor activity, haematopoiesis regulating  
 C activity, tissue growth factor activity, immunomodulatory activity and  
 C activin/inhibin activity and may be useful in the diagnosis and/or  
 C treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 C inflammation. Note: The sequence data for this patent did not form part  
 C of the printed specification, but was obtained in electronic format  
 C directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 X C  
 X C Sequence 33 AA;  
 X C  
 X C Query Match 100.0%; Score 19; DB 4; Length 33;  
 X C Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
 X C Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 X C  
 X C 1 VAEF 4  
 X C 3 VAEF 6  
 X C  
 X C RESULT 33  
 X C AR77953  
 X C D AAR77953 standard; peptide; 35 AA.  
 X C X AAR77953;  
 X C T 09-OCT-1996 (first entry)  
 X C X Antigenic Tbp2 peptide TBP2-21.  
 X C E Tbp1; Tbp2; transferrin receptor operon; vaccine; antigen;  
 X C W non-typable strain; Haemophilus influenzae; meningitis.  
 X C S Synthetic.  
 X C N WO9513370-A1.  
 X C X 18-MAY-1995.  
 X C D 07-NOV-1994; 94WO-CA000616.  
 X C F 08-NOV-1993; 93US-00148968.  
 X C R 29-DEC-1993; 93US-00175116.  
 X C X (CONN-) CONNAUGHT LAB LTD.  
 X C A Loosmore S, Harkness R, Schryvers A, Chong P, Gray-Owen S;  
 X C I Yang Y, Mordin A, Klein M;  
 X C X WPI; 1995-194089/25.  
 X C R Nucleic acids encoding Haemophilus transferrin receptor - used to develop  
 X C T prods for detection and in diagnosis, prevention and treatment of  
 X C T Haemophilus infection.  
 X C S Example 16; Page 72; 231pp; English.  
 X C X AAR77933-969 are predicted antigenic peptides derived from conserved  
 X C C regions of the Tbp2 protein from H. influenzae strains Egan, Minna, DL63  
 X C C and non-typable strain PA12085. The transferrin receptor (TfR) operon  
 X C C consists of two genes (Tbp1 and Tbp2) arranged in tandem and which are  
 X C C transcribed from a single promoter. H. influenzae TfR is iron- and/or  
 X C C haemin-regulated and a putative fur-binding site has been identified  
 X C C upstream of Tbp2. Antibodies blocking this binding site may prevent  
 X C C bacterial growth. Fragments of the TfR (or its genes) are useful in  
 X C C vaccines to provide protection against, e.g. bacterial meningitis. An  
 X C C advantage of using the TfR is that it shares homology with TfR of other

CC H. influenzae strains including non-typable strains. According to the  
 CC specification the present sequence shows residues 449-484 of Tbp2 from  
 CC the H. influenzae strain Egan  
 XX Sequence 35 AA;  
 SQ  
 Query Match 100.0%; Score 19; DB 2; Length 35;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 DB 14 VAEF 17  
 RESULT 34  
 AAW46146  
 ID AAW46146 standard; protein; 35 AA.  
 XX  
 AC AAW46146;  
 XX  
 DT 05-MAY-1998 (first entry)  
 XX  
 DE Predicted antigenic Tbp1 peptide TBP2-21.  
 XX  
 XW Transferrin receptor; Haemophilus influenzae type b; iron;  
 XW human transferrin; iron source; antibody; bacterial growth; vaccine;  
 XW immunogenic truncated analogue; antigen; Tbp1; Tbp2.  
 XX  
 OS Synthetic.  
 OS Haemophilus influenzae.  
 PN WO9640929-A2.  
 XX  
 PD 19-DEC-1996.  
 XX  
 PF 07-JUN-1996; 96WO-CA000399.  
 XX  
 PR 07-JUN-1995; 95US-00483577.  
 PR 17-MAY-1996; 96US-00649518.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Loosmore SM, Harkness RB, Schryvers AB, Chong P, Gray-Owen S;  
 PI Yang Y, Mordin AD, Klein MH;  
 DR WPI; 1997-052329/05.  
 XX  
 XX Haemophilus truncated transferrin receptor protein analogue, Tbp2 - used  
 XX to induce protection against disease caused by transferrin-producing  
 XX pathogens, or as antigen to detect Haemophilus Tfr antibodies.  
 PS Example 16; Page 70; 228pp; English.  
 XX  
 CC AAW46126-62 are predicted antigenic peptides derived from the Tbp2  
 CC protein of Haemophilus influenzae type b. Tbp2 is part of the transferrin  
 CC receptor, of which Tbp1 is also a subunit. The deduced amino acid  
 CC sequences of Tbp1 and Tbp2 were compared, and regions of conservation  
 CC identified. The above peptides are derived from these regions, the  
 CC present peptide being derived from residues 449-484. Iron is an essential  
 CC nutrient for the growth of these bacteria, and they can utilise human  
 CC transferrin as a source of iron. Antibodies which block the access of the  
 CC transferrin receptor to its iron source prevent bacterial growth. The  
 CC transferrin receptor, or fragments, therefore, are good vaccine  
 CC candidates. An immunogenic composition comprising (or encoding) the  
 CC immunogenic truncated analogue can be used to induce protection against a  
 CC disease caused by a bacterial pathogen that produces the transferrin  
 CC receptor. The immunogenic truncated analogue is also useful as an antigen  
 CC in immunoassays for the detection of Haemophilus transferrin receptor  
 CC antibodies, while the nucleic acid molecule can be used as a  
 CC hybridisation probe for the detection of other transferrin receptor genes  
 XX  
 SQ Sequence 35 AA;

Query Match 100.0%; Score 19; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 14 VAEF 17

RESULT 35  
AAW51751  
ID AAY51751 standard; protein; 35 AA.

XX AC AAY51751;  
XX DT 13-JUN-2000 (first entry)  
XX DE H. influenzae antigenic Tbp2 peptide TBP2-21.  
XX KW Transferrin receptor; Tbp1; Tbp2; immunogenic; antibacterial; vaccine;  
XX diagnosis.  
XX OS Haemophilus influenzae.  
XX PN US6015688-A.  
XX PD 18-JAN-2000.  
XX PF 07-JUN-1995; 95US-00483577.  
XX PR 08-NOV-1993; 93US-00148968.  
XX PR 29-DEC-1993; 93US-00175116.  
XX PR 08-NOV-1994; 94US-00337483.  
XX PA (CONN-) CONNAUGHT LAB LTD.  
XX PI Loomore S, Harkness R, Schryvers A, Gray-Owen S, Yang Y;  
XX PI Murdin A, Klein M, Chong P;  
XX WPI; 1997-052329/05.

XX Haemophilus truncated transferrin receptor protein analogue, Tbp2 - used  
XX to induce protection against disease caused by transferrin producing  
XX pathogens, or as antigen to detect Haemophilus Tfr antibodies.  
XX Example 16; Col 39-40; 281pp; English.

XX This invention describes a novel isolated and purified nucleic acid (I)  
XX encoding an immunogenic, C-terminally truncated analog of one of the  
XX transferrin receptor proteins Tbp1 or Tbp2 of Haemophilus influenzae  
XX which has antibacterial activity. (I) are used for recombinant production  
XX of truncated Tbp; as probes and primers for detecting, and diagnosing  
XX infection by, Haemophilus, also for isolating similar sequences from  
XX other bacteria; as immunogens for vaccinating against infections caused  
XX by bacteria that produce transferrin receptors, e.g. Haemophilus,  
XX Neisseria or Branhamella. The truncated proteins are useful as immunogens  
XX (as above); for diagnosing infection (as antigens in immunoassays) and  
XX for raising antibodies, used for diagnosis of infections or for passive  
XX immunization. AAY51695-Y51767 represent H. influenzae transferrin  
XX receptor proteins Tbp1 and Tbp2 antigenic peptide fragments

XX SQ Sequence 35 AA;

Query Match 100.0%; Score 19; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 14 VAEF 17

RESULT 36  
AAW55788  
ID AAW55788 standard; peptide; 35 AA.  
XX AC AAW55788;  
XX DT 14-JUL-1998 (first entry)  
XX DE Human aquaporin-1 NH2-terminal amino acid sequence CHIP28.  
XX KW Human; aquaporin-1; AQP-1; water channel protein; regulation;  
XX osmotic change; erythrocyte; dryness; blindness; hydration; asthma;  
XX secretion.  
XX OS Homo sapiens.  
XX FT Key Location/Qualifiers  
XX FT Misc-difference 30 /label= Unknown  
XX FT /note= "not specified but is given as Ser in the full  
XX FT length protein given in AAW55786"

XX PN US5741671-A.

XX PD 21-APR-1998.

XX PF 06-JUN-1995; 95US-00468763.

XX PR 17-AUG-1992; 92US-00930168.

XX PR 24-FEB-1995; 95US-00393996.

XX PA (UYJO ) UNIV JOHNS HOPKINS.

XX PI Agre PC;

XX WPI; 1998-260501/23.

XX Polynucleotide(s) encoding water channel protein Aquaporin-1 - are useful  
XX for recombinant production of protein for activity studies.

XX Example 2; Col 20; 48pp; English.

XX The present sequence represents the NH2-terminal amino acid sequence of  
XX aquaporin-1 (AQP-1), isolated from human erythrocytes. AQP-1 is a water  
XX channel protein. Water channel proteins regulate the passage of water in  
XX and out of cells, in response to osmotic changes. The DNA encoding AQP-1  
XX is useful for the recombinant production of AQP-1, found in mammalian  
XX erythrocytes, and is useful in the study to identify reagents which  
XX enhance or inhibit water channel function. This can lead to therapeutics  
XX which enhance secretion e.g. in the case of dryness of eyes which can  
XX lead to blindness or to hydrate large respiratory airways, as their  
XX dryness can precipitate asthma

XX SQ Sequence 35 AA;

Query Match 100.0%; Score 19; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 14 VAEF 17

RESULT 37  
AAW54105  
ID AAW54105 standard; peptide; 35 AA.

XX AC AAW54105;

XX DT 20-JUL-1998 (first entry)

XX DE Tbp2 antigenic peptide TBP2-21.

M tbp1; tbp2; vaccine; H. influenzae; antibody; diagnosis;  
 X passive immunisation; transferrin receptor operon.  
 X Haemophilus influenzae.  
 X US5708149-A.  
 X 13-JAN-1998.  
 X 07-JUN-1995; 95US-00487890.  
 X 08-NOV-1993; 93US-00148968.  
 X 29-DEC-1993; 93US-00175116.  
 X 08-NOV-1994; 94US-00337483.  
 X (CONN-) CONNAUGHT LAB LTD.  
 X Gray-Owen S, Klein M, Harkness R, Loosmore S, Yang Y, Chong P;  
 X Murdin A, Schryvers A;  
 X WPI; 1998-100410/09.  
 X Purification of recombinant Haemophilus transferrin-binding protein - by  
 X solubilising inclusion bodies separated from cell lysate.  
 X Example 16/17; Column 37-38; 261pp; English.  
 X Peptides AAW53086-W53100 and peptides AAW54100-W54122 are derived from  
 X the Tbp2 protein. The Tbp2 protein is one of two proteins with genes  
 X found on the transferrin operon. These peptides can be used along with  
 X the genes, DNA sequences and recombinant proteins for diagnosis,  
 X immunisation and the generation of diagnostic and immunological reagents.  
 X They can also be used to protect from bacteria that produce transferrin  
 X receptor protein  
 X Sequence 35 AA;  
 X  
 X Query Match 100.0%; Score 19; DB 2; Length 35;  
 X Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 X Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 X  
 X 1 VAEP 4  
 X ||||  
 X 14 VAEP 17  
 X  
 X RESULT 38  
 X W94320  
 X AAW94320 standard; peptide; 35 AA.  
 X AAW94320;  
 X 12-APR-1999 (first entry)  
 X Human aquaporin-1 (CHIP28) N-terminal peptide.  
 X Rat; aquaporin-5; AQP5; AQP1; transmembrane water channel protein;  
 X major intrinsic protein; MIP; CHIP28.  
 X Homo sapiens.  
 X Key Location/Qualifiers  
 X Misc-difference 30  
 X /note= "unspecified"  
 X US5858702-A.  
 X 12-JAN-1999.  
 X 24-FEB-1995; 95US-00393996.  
 X 13-DEC-1991; 91US-00806273.

PR 17-AUG-1992; 92US-00930168.  
 XX (UYJO ) UNIV JOHNS HOPKINS.  
 XX PI  
 XX Agre PC;  
 XX WPI; 1999-152100/13.  
 XX DNA encoding aquaporin-5 water channel protein - useful for producing  
 PT recombinant protein.  
 XX Example 2; Col 20; 47pp; English.  
 XX The present invention describes a water channel protein called aquaporin-  
 CC 5 (AQP5). The polynucleotide encoding AQP5 is useful for producing  
 CC recombinant AQP5, which can be incorporated into proteoliposomes or cell  
 CC membrane vesicles, which are able to be used in screening assays for water  
 CC channel agonists or antagonists. The present sequence represents the N-  
 CC terminal peptide from AQP1 (also called CHIP28), from an example of the  
 CC present invention  
 XX Sequence 35 AA;  
 X  
 X Query Match 100.0%; Score 19; DB 2; Length 35;  
 X Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 X Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 X  
 X 1 VAEP 4  
 X ||||  
 X 14 VAEP 17  
 X  
 X RESULT 39  
 X AAY80448  
 X ID AAY80448 standard; peptide; 35 AA.  
 X AC AAY80448;  
 X DT 06-JUN-2000 (first entry)  
 X H. influenzae transferrin receptor Tbp1 epitope TBP2-21.  
 X Antibacterial; antiinflammatory; auditory; respiratory; antibody;  
 X antiserum; transferrin receptor; immunogen; epitope; otitis media;  
 X bacterial meningitis; epiglottitis; pneumonia; tracheobronchitis.  
 X Haemophilus influenzae.  
 X US6008326-A.  
 X 28-DEC-1999.  
 X 07-JUN-1995; 95US-00474671.  
 X 08-NOV-1993; 93US-00148968.  
 X 29-DEC-1993; 93US-00175116.  
 X 08-NOV-1995; 95US-00337483.  
 X (CONN-) CONNAUGHT LAB LTD.  
 X Loosmore S, Harkness R, Chong P, Gray-Owen S, Yang Y, Klein M;  
 X Murdin A, Schryvers A;  
 X WPI; 2000-096387/08.  
 X Antibodies specific for transferrin receptor proteins of Haemophilus  
 PT influenzae, useful for treating otitis media, epiglottitis, pneumonia and  
 PT tracheobronchitis.  
 XX Disclosure; Col 37-38; 252pp; English.  
 XX The invention relates to novel antibodies (or monospecific antisera)  
 CC specific for single transferrin receptor proteins (or immunogenic

CC fragment) from strains of *Haemophilus influenzae*. This sequence  
 CC corresponds to an epitope from the *H. influenzae* transferrin receptor  
 CC protein Tbp2. The antibodies may be used for preventing and treating  
 CC infections and disorders caused by *H. influenzae*, including bacterial  
 CC meningitis, otitis media, epiglottitis, pneumonia and tracheobronchitis.  
 CC The antibodies may also be used to detect the presence of *H. influenzae*  
 CC proteins in samples according to standard methodologies (e.g. enzyme  
 CC linked immunosorbent assay (ELISA)) and hence diagnose infections

XX Sequence 35 AA;  
 SQ Query Match 100.0%; Score 19; DB 3; Length 35;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VAEF 4  
 |||||  
 Db 14 VAEF 17

RESULT 40  
 AAW27782  
 ID AAW27782 standard; protein; 36 AA.

XX AC AAW27782;  
 XX DT 21-JUL-1998 (first entry)  
 XX DE UDP-N-acetylglucosamine 1-carboxyvinyltransferase.

XX KW *Staphylococcus aureus* protein; ribozyme; antisense sequence; control;  
 KW *Staphylococcus aureus* gene; regulatory element; bacterial gene expression;  
 KW vaccine; *Staphylococcus aureus* infection; food poisoning; scaled skin syndrome;  
 KW toxic shock syndrome.

XX OS *Staphylococcus aureus*.

XX FH Key Location/Qualifiers  
 FT Misc-difference 34  
 FT /note= "not specified"

XX WO9730070-A1.

XX PD 21-AUG-1997.

XX PF 19-FEB-1997; 97WO-US002318.

XX PR 20-FEB-1996; 96US-0011888P.

XX PA (SMK ) SMITHKLINE BEECHAM CORP.

XX PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;  
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX DR WPI; 1997-424969/39.

XX DR N-PSDB; AAT83751.

XX PT Novel polypeptide(s) from *Staphylococcus aureus* strain WCUH29 - used to  
 PT isolate antimicrobial compounds, and in vaccines against *S. aureus*  
 PT infection.

XX PS Claim 6; Page 272; 989pp; English.

XX CC The present sequence represents a *Staphylococcus aureus* protein, that,  
 CC based on homology with a *Bacillus subtilis* protein, is believed to be a  
 CC probable UDP-N-acetylglucosamine 1-carboxyvinyltransferase (enoylpyruvate  
 CC transferase, UDP-N-acetylglucosamine enoylpyruvyl transferase). The DNA  
 CC sequence was isolated from a library of clones of *S. aureus* WCUH 29 in  
 CC *Escherichia coli*. The DNA sequences can be used in the construction of  
 CC ribozymes and antisense sequences to control the expression of  
 CC *Staphylococcus aureus* genes. The DNA sequence is also useful as a source of  
 CC regulatory elements for the control of bacterial gene expression. The  
 CC present protein may be used to produce vaccines to enable a host to

CC produce specific antibodies with antibacterial action. These vaccines and  
 CC antibodies would protect a host against invasion by *S. aureus*, and  
 CC conditions relating to *Staphylococcus aureus* infection, e.g. *Staphylococcus*  
 CC poisoning, scaled skin syndrome, and toxic shock syndrome

SQ Sequence 36 AA;  
 Query Match 100.0%; Score 19; DB 2; Length 36;  
 Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VAEF 4  
 |||||  
 Db 3 VAEF 6

RESULT 41  
 AAM99833  
 ID AAM99833 standard; protein; 37 AA.

XX AC AAM99833;

XX DT 07-JAN-2002 (first entry)

XX DE Human excretory related polypeptide SEQ ID NO 570.

XX KW Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
 KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;  
 KW antineumatic; hepatotropic; cerebroprotective; antinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine;  
 KW excretory system.

XX OS *Homo sapiens*.

XX PN WO200155313-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001323.

XX PR 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184664P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

XX PR 17-MAR-2000; 2000US-0190076P.

XX PR 18-APR-2000; 2000US-0198123P.

XX PR 19-MAY-2000; 2000US-0205515P.

XX PR 07-JUN-2000; 2000US-0209467P.

XX PR 28-JUN-2000; 2000US-0214886P.

XX PR 30-JUN-2000; 2000US-0215135P.

XX PR 07-JUL-2000; 2000US-0216647P.

XX PR 07-JUL-2000; 2000US-0216880P.

XX PR 11-JUL-2000; 2000US-0217487P.

XX PR 11-JUL-2000; 2000US-0217486P.

XX PR 14-JUL-2000; 2000US-0218290P.

XX PR 26-JUL-2000; 2000US-0220963P.

XX PR 26-JUL-2000; 2000US-0220964P.

XX PR 14-AUG-2000; 2000US-0224518P.

XX PR 14-AUG-2000; 2000US-0224519P.

XX PR 14-AUG-2000; 2000US-0225213P.

XX PR 14-AUG-2000; 2000US-0225214P.

XX PR 14-AUG-2000; 2000US-0225266P.

XX PR 14-AUG-2000; 2000US-0225267P.

XX PR 14-AUG-2000; 2000US-0225268P.

XX PR 14-AUG-2000; 2000US-0225270P.

XX PR 14-AUG-2000; 2000US-0225447P.

XX PR 14-AUG-2000; 2000US-0225757P.

XX PR 14-AUG-2000; 2000US-0225758P.

XX PR 14-AUG-2000; 2000US-0225759P.

R 18-AUG-2000; 2000US-0226279P.  
R 22-AUG-2000; 2000US-0226681P.  
R 22-AUG-2000; 2000US-0226688P.  
R 22-AUG-2000; 2000US-0227182P.  
R 23-AUG-2000; 2000US-0227009P.  
R 30-AUG-2000; 2000US-0228924P.  
R 01-SEP-2000; 2000US-0229287P.  
R 01-SEP-2000; 2000US-0229343P.  
R 01-SEP-2000; 2000US-0229344P.  
R 01-SEP-2000; 2000US-0229345P.  
R 05-SEP-2000; 2000US-0229509P.  
R 05-SEP-2000; 2000US-0229513P.  
R 05-SEP-2000; 2000US-0230437P.  
R 06-SEP-2000; 2000US-0230438P.  
R 08-SEP-2000; 2000US-0231242P.  
R 08-SEP-2000; 2000US-0231243P.  
R 08-SEP-2000; 2000US-0231244P.  
R 08-SEP-2000; 2000US-0231413P.  
R 08-SEP-2000; 2000US-0231414P.  
R 08-SEP-2000; 2000US-0232080P.  
R 08-SEP-2000; 2000US-0232081P.  
R 12-SEP-2000; 2000US-0231968P.  
R 14-SEP-2000; 2000US-0232397P.  
R 14-SEP-2000; 2000US-0232398P.  
R 14-SEP-2000; 2000US-0232399P.  
R 14-SEP-2000; 2000US-0232400P.  
R 14-SEP-2000; 2000US-0232401P.  
R 14-SEP-2000; 2000US-0232633P.  
R 14-SEP-2000; 2000US-0233064P.  
R 14-SEP-2000; 2000US-0233065P.  
R 21-SEP-2000; 2000US-0234223P.  
R 21-SEP-2000; 2000US-0234274P.  
R 23-SEP-2000; 2000US-0234597P.  
R 23-SEP-2000; 2000US-0234598P.  
R 26-SEP-2000; 2000US-0235484P.  
R 27-SEP-2000; 2000US-0235834P.  
R 27-SEP-2000; 2000US-0235836P.  
R 29-SEP-2000; 2000US-0236327P.  
R 29-SEP-2000; 2000US-0236367P.  
R 29-SEP-2000; 2000US-0236368P.  
R 29-SEP-2000; 2000US-0236369P.  
R 29-SEP-2000; 2000US-0236370P.  
R 02-OCT-2000; 2000US-0236802P.  
R 02-OCT-2000; 2000US-0237037P.  
R 02-OCT-2000; 2000US-0237038P.  
R 02-OCT-2000; 2000US-0237039P.  
R 02-OCT-2000; 2000US-0237040P.  
R 13-OCT-2000; 2000US-0239335P.  
R 13-OCT-2000; 2000US-0239337P.  
R 20-OCT-2000; 2000US-0240960P.  
R 20-OCT-2000; 2000US-0241221P.  
R 20-OCT-2000; 2000US-0241785P.  
R 20-OCT-2000; 2000US-0241786P.  
R 20-OCT-2000; 2000US-0241787P.  
R 20-OCT-2000; 2000US-0241808P.  
R 20-OCT-2000; 2000US-0241809P.  
R 20-OCT-2000; 2000US-0241826P.  
R 01-NOV-2000; 2000US-0244617P.  
R 08-NOV-2000; 2000US-0246474P.  
R 08-NOV-2000; 2000US-0246475P.  
R 08-NOV-2000; 2000US-0246476P.  
R 08-NOV-2000; 2000US-0246477P.  
R 08-NOV-2000; 2000US-0246478P.  
R 08-NOV-2000; 2000US-0246523P.  
R 08-NOV-2000; 2000US-0246524P.  
R 08-NOV-2000; 2000US-0246525P.  
R 08-NOV-2000; 2000US-0246526P.  
R 08-NOV-2000; 2000US-0246527P.  
R 08-NOV-2000; 2000US-0246528P.  
R 08-NOV-2000; 2000US-0246532P.  
R 08-NOV-2000; 2000US-0246609P.  
R 08-NOV-2000; 2000US-0246610P.  
R 08-NOV-2000; 2000US-0246611P.

PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251858P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-02559678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465569/50.

N-PSDB; AA198806.

Isolated nucleic acid molecule encoding excretory system antigen is used in preventing, treating or ameliorating a medical condition.

Claim 11; SEQ ID NO 570; 574pp + Sequence Listing; English.

The invention relates to novel excretory system related human polynucleotides (AA198567-AA199503) and the encoded proteins (AA99594-AA99913) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy, especially disorders related to the excretory system. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 37 AA;

Query Match 100.0%; Score 19; DB 4; Length 37;

Best Local Similarity 100.0%; Pred. No. 5e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4

QY

PR	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
PR	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

R 17-NOV-2000; 2000US-0249244P.  
 R 17-NOV-2000; 2000US-0249245P.  
 R 17-NOV-2000; 2000US-0249264P.  
 R 17-NOV-2000; 2000US-0249265P.  
 R 17-NOV-2000; 2000US-0249297P.  
 R 17-NOV-2000; 2000US-0249299P.  
 R 17-NOV-2000; 2000US-0249300P.  
 R 01-DEC-2000; 2000US-0250160P.  
 R 01-DEC-2000; 2000US-0250391P.  
 R 05-DEC-2000; 2000US-0251030P.  
 R 05-DEC-2000; 2000US-0251388P.  
 R 05-DEC-2000; 2000US-0256719P.  
 R 06-DEC-2000; 2000US-0251479P.  
 R 08-DEC-2000; 2000US-0251856P.  
 R 08-DEC-2000; 2000US-0251868P.  
 R 08-DEC-2000; 2000US-0251869P.  
 R 08-DEC-2000; 2000US-0251989P.  
 R 08-DEC-2000; 2000US-0251990P.  
 R 11-DEC-2000; 2000US-0254097P.  
 R 05-JAN-2001; 2001US-0259678P.  
 X (HUMA-) HUMAN GENOME SCI INC.  
 X  
 X Rosen CA, Barash SC, Ruben SM;  
 I  
 X WPI: 2001-488784/53.  
 R  
 R N-PSDB; AAI63202.  
 X

T New isolated nucleic acids and polypeptides, useful for diagnosing,  
 T treating and/or preventing human diseases and disorders.  
 T  
 X

S Claim 11; SEQ ID NO 517; 564pp + Sequence Listing; English.

X The invention relates to novel kidney related polynucleotides (AAI62971-  
 X AAI63793) and the encoded polypeptides (AAI62417-AAI63691) collectively  
 C known as kidney antigens and the use of such kidney antigens for  
 C detecting disorders of the kidney, especially kidney cancer and kidney  
 C cancer metastases. The polynucleotides and proteins are also useful for  
 C preventing, treating or ameliorating medical conditions e.g. by protein  
 C or gene therapy. The genes are isolated from a range of human tissues  
 C disclosed in the specification. The nucleic acids, proteins, antibodies  
 C and (ant)agonists are useful in the diagnosis, treatment and prevention  
 C of: (a) cancer, e.g. breast and ovarian cancer, and other cancers of the  
 C adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
 C lung, or urogenital; (b) immune disorders e.g. Addison's disease,  
 C allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
 C diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
 C arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
 C myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.  
 C cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
 C bacterial, fungal and parasitic infections. Note: The sequence data for  
 C this patent did not form part of the printed specification, but was  
 C obtained in electronic format directly from WIPO at  
 C ftp.wipo.int/pub/published\_pct\_sequences

X Sequence 37 AA;

Query Match 100.0%; Score 19; DB 4; Length 37;  
 Best Local Similarity 100.0%; Pred. No. 5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 14 VAEF 17

RESULT 43

AAI17544

D AAI17544 standard; protein; 43 AA.

X AAI17544;

X AAI17544;

Y 12-OCT-2001 (first entry)

XX Peptide #3978 encoded by probe for measuring cervical gene expression.  
 DE Probe; human; microarray; gene expression; cervical epithelial cell;  
 XX cervical cancer.  
 KW Homo sapiens.  
 XX WO200157278-A2.  
 DN  
 XX 09-AUG-2001.  
 XX  
 XX 30-JAN-2001; 2001WO-US000670.  
 XX  
 XX 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 01-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 XX (WOLB-) MOLECULAR DYNAMICS INC.  
 PA  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI  
 XX WPI: 2001-488901/53.  
 DR  
 XX

PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human cervical epithelial cells.  
 PT

PS Claim 27; SEQ ID NO 22370; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes  
 CC (SENPs: see AAI10068-AAI28459). The present sequence is a peptide encoded  
 CC by one such probe. The SENPs are derived from human Hela cells. The SENPs  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably  
 CC cervical cancer. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 43 AA;

Query Match 100.0%; Score 19; DB 4; Length 43;

Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAEF 4  
 40 VAEF 43

RESULT 44

ABB31362

ID ABB31362 standard; peptide; 43 AA.

XX ABB31362;

AC ABB31362;

DT 01-FEB-2002 (first entry)

DE Peptide #4013 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast; disease;  
 KW cancer.  
 XX Homo sapiens.

OS

XX WO200157271-A2.

PN

XX 09-AUG-2001.

PD

XX PF 30-JAN-2001; 2001WO-US000662.  
 XX PR 04-FEB-2000; 2000US-0180312P.  
 XX PR 26-MAY-2000; 2000US-0207456P.  
 XX PR 30-JUN-2000; 2000US-00608408.  
 XX PR 03-AUG-2000; 2000US-00632366.  
 XX PR 21-SEP-2000; 2000US-0234687P.  
 XX PR 27-SEP-2000; 2000US-0236359P.  
 XX PR 04-OCT-2000; 2000GB-00024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX PI WPI; 2001-496933/54.  
 XX DR  
 XX DR WPI; 2001-496933/54.  
 XX PT New spatially-addressable set of single exon nucleic acid probes, useful  
 XX PT for measuring gene expression in sample derived from human breast,  
 XX PT comprises number of single exon nucleic acid probes.  
 XX PS Claim 27; SEQ ID NO 14330; 327pp + Sequence Listing; English.  
 XX CC The invention relates to a spatially-addressable set of single exon  
 XX CC nucleic acid probes for measuring gene expression in a sample derived  
 XX CC from human breast and BT 474 cells. The method involves contacting the  
 XX CC probes with a collection of detectably labelled nucleic acids derived  
 XX CC from mRNA of human breast, and then measuring the label bound to each  
 XX CC probe of the microarray. The probes are useful for verifying the  
 XX CC expression of regions of genomic DNA predicted to encode proteins. They  
 XX CC are useful for gene discovery, and for determining predisposition and/or  
 XX CC assessing the toxicity of chemical agents on cells. The microarray of  
 XX CC this invention presents a far greater diversity of probes for measuring  
 XX CC gene expression, with far less bias than expressed sequence tag  
 XX CC microarrays. The method is suitable for rapid production of functional  
 XX CC information from genomic sequence. The present sequence is a peptide  
 XX CC encoded by a single exon nucleic acid probe of the invention. Note: The  
 XX CC sequence data for this patent did not form part of the printed  
 XX CC specification, but was obtained in electronic format directly from WIPO  
 XX CC at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 43 AA;  
 Query Match 100.0%; Score 19; DB 4; Length 43;  
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 Db 40 VAEF 43  
 |||||

RESULT 45  
 ABB21904  
 ID ABB21904 standard; protein; 43 AA.  
 XX AC ABB21904;  
 XX DT 23-JAN-2002 (first entry)  
 XX DE Protein #3903 encoded by probe for measuring heart cell gene expression.  
 XX KW Human; gene expression; heart; microarray; vascular system;  
 XX KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 XX KW congenital heart disease.  
 XX OS Homo sapiens.  
 XX XX WO200157274-A2.  
 XX PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000666.  
 XX PR 04-FEB-2000; 2000US-0180312P.  
 XX PR 26-MAY-2000; 2000US-0207456P.  
 XX PR 30-JUN-2000; 2000US-00608408.  
 XX PR 03-AUG-2000; 2000US-00632366.  
 XX PR 21-SEP-2000; 2000US-0234687P.  
 XX PR 27-SEP-2000; 2000US-0236359P.  
 XX PR 04-OCT-2000; 2000GB-00024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX PI WPI; 2001-488899/53.  
 XX DR  
 XX DR WPI; 2001-488899/53.  
 XX PT Single exon nucleic acid probes for analyzing gene expression in human  
 XX PT hearts.  
 XX PS Claim 15; SEQ ID NO 23674; 530pp; English.  
 XX CC The present invention relates to single exon nucleic acid probes for  
 XX CC measuring human gene expression in a sample derived from human heart (see  
 XX CC ABA1535-ABA41305). The present sequence is a protein encoded by one such  
 XX CC probe. The probes may be used for predicting, measuring and displaying  
 XX CC gene expression in samples derived from the human heart via microarrays.  
 XX CC By measuring gene expression, the probes are useful for predicting,  
 XX CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 XX CC human heart and vascular system e.g. cardiovascular disease,  
 XX CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
 XX CC sequence data for this patent did not form part of the printed  
 XX CC specification, but was obtained in electronic format directly from WIPO  
 XX CC at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 43 AA;  
 Query Match 100.0%; Score 19; DB 4; Length 43;  
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 Db 40 VAEF 43  
 |||||

RESULT 46  
 ABG51414  
 ID ABG51414 standard; peptide; 43 AA.  
 XX AC ABG51414;  
 XX DT 25-FEB-2003 (first entry)  
 XX DE Human liver peptide, SEQ ID No 30062.  
 XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 XX KW hypercholesterolaemia; coronary heart disease.  
 XX OS Homo sapiens.  
 XX XX WO200157273-A2.  
 XX PD 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000664.  
 XX PR 04-FEB-2000; 2000US-0180312P.  
 XX PR 26-MAY-2000; 2000US-0207456P.  
 XX PR 30-JUN-2000; 2000US-00608408.  
 XX PR 03-AUG-2000; 2000US-00632366.  
 XX PR 21-SEP-2000; 2000US-0234687P.  
 XX PR 27-SEP-2000; 2000US-0236359P.  
 XX PR 04-OCT-2000; 2000GB-00024263.

X (MOLE-) MOLECULAR DYNAMICS INC.  
 X Penn SG, Hanzel DK, Chen W, Rank DR;  
 X WPI; 2001-488998/53.  
 X Human genome-derived single exon nucleic acid probes useful for analyzing  
 X gene expression in human adult liver.  
 S Claim 27; SEQ ID NO 30062; 658pp; English.  
 C The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 C measuring human gene expression in a sample derived from human adult  
 C liver, comprising one of 13109 defined nucleotide sequences given in the  
 C specification (or complements/ fragments). The probe hybridises at high  
 C stringency to a nucleic acid molecule expressed in the human adult liver.  
 C (I) may be used for predicting, measuring and displaying gene expression  
 C in samples derived from human adult liver. The genes identified may be  
 C involved in genetic liver diseases such as cirrhosis,  
 C hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 C associated with coronary heart disease. ABG47348-ABG59930 represent human  
 C liver single exon encoded peptides of the invention. Note: The sequence  
 C information for this patent does not appear in the printed specification  
 C but was obtained in electronic format directly from WIPO at  
 C ftp.wipo.int/pub/published\_pct\_sequences  
 X Sequence 43 AA;  
 Query Match 100.0%; Score 19; DB 4; Length 43;  
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 1 VAEF 4  
 b ||||  
 40 VAEF 43  
 RESULT 47  
 AM05214  
 D AAM05214 standard; protein; 43 AA.  
 X AAM05214;  
 X 09-OCT-2001 (first entry)  
 X Peptide #3896 encoded by probe for measuring breast gene expression.  
 X Probe; human; breast disease; breast cancer; developmental disorder;  
 X inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
 X Homo sapiens.  
 X WO200157270-A2.  
 X 09-AUG-2001.  
 X 29-JAN-2001; 2001WO-US000661.  
 X 04-FEB-2000; 2000US-0180312P.  
 X 26-MAY-2000; 2000US-0207456P.  
 X 30-JUN-2000; 2000US-00608408.  
 X 03-AUG-2000; 2000US-00632366.  
 X 21-SEP-2000; 2000US-0234687P.  
 X 27-SEP-2000; 2000US-0236359P.  
 X 04-OCT-2000; 2000GB-00024263.  
 X (MOLE-) MOLECULAR DYNAMICS INC.  
 X Penn SG, Hanzel DK, Chen W, Rank DR;  
 X WPI; 2001-476286/51.

PT Novel single exon nucleic acid probe used to measuring gene expression in  
 PT a human breast.  
 XX Claim 27; SEQ ID NO 13954; 322pp; English.  
 PS  
 XX The present invention relates to novel single exon nucleic acid probes  
 CC (see AA10010-AA110057). The present sequence is a peptide encoded by one  
 CC a human breast sample, where the probe hybridises at high stringency to a  
 CC nucleic acid expressed in the human breast. The probes are useful for  
 CC predicting, diagnosing, grading, staging, monitoring and prognosing  
 CC diseases of the human breast, particularly those diseases with polygenic  
 CC aetiology. The diseases include: breast cancer, disorders of development,  
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative,  
 CC breast disease and non-carcinoma tumours. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 43 AA;  
 SQ  
 Query Match 100.0%; Score 19; DB 4; Length 43;  
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 Db ||||  
 40 VAEF 43  
 RESULT 48  
 AAB27178  
 ID AAB27178 standard; protein; 47 AA.  
 XX AAB27178;  
 AC AAB27178;  
 XX 27-FEB-2001 (first entry)  
 XX RSV partial protein sequence SEQ ID NO: 26.  
 XX Negative stranded RNA virus; vaccine; attenuated virus; RSV; PIV;  
 XX measles; respiratory syncytial virus; parainfluenza virus.  
 XX Respiratory syncytial virus.  
 XX WO200061737-A2.  
 XX 19-OCT-2000.  
 XX 12-APR-2000; 2000WO-US009695.  
 XX 13-APR-1999; 99US-0129006P.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Murphy BR, Collins PL, Durbin AP, Skiadopoulos ME;  
 XX WPI; 2000-687044/67.  
 XX Producing attenuated negative stranded RNA virus vaccines from cloned  
 PT sequences, useful for immunizing against e.g. respiratory syncytial  
 PT virus, human parainfluenza virus, Sendai virus Newcastle disease virus,  
 PT mumps virus and measles virus.  
 XX Example 1; Page 62; 137pp; English.  
 XX The present invention is concerned with producing vaccines against  
 CC negative stranded RNA viruses. These viruses include measles, respiratory  
 CC syncytial virus (RSV) and parainfluenza virus (PIV) in particular. The  
 CC method of the invention comprises the production of a mutated form of the  
 CC virus which attenuates the strain and enables it to be used as a vaccine.  
 CC The present sequence comprises a partial viral protein sequence

XX	Query Match	100.0%; Score 19; DB 3; Length 47;
XX	Best Local Similarity	100.0%; Pred. No. 6.4e+02;
XX	Mismatches	0; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 VAEF 4	
DB	34 VAEF 37	
DE		
DE	12-OCT-2001 (first entry)	
DE	Peptide #4975 encoded by probe for measuring cervical gene expression.	
KW	Probe; human; microarray; gene expression; cervical epithelial cell;	
KW	cervical cancer.	
OS	Homo sapiens.	
PN	WO200157278-A2.	
XX	09-AUG-2001.	
XX	30-JAN-2001; 2001WO-US000670.	
PR	04-FEB-2000; 2000US-0180312P.	
PR	26-MAY-2000; 2000US-0207456P.	
PR	30-JUN-2000; 2000US-00608408.	
PR	03-AUG-2000; 2000US-00632366.	
PR	21-SEP-2000; 2000US-0234687P.	
PR	27-SEP-2000; 2000US-0236359P.	
PR	04-OCT-2000; 2000GB-00024263.	
XX	(MOLE-) MOLECULAR DYNAMICS INC.	
XX	Penn SG, Hanzel DK, Chen W, Rank DR;	
XX	WPI; 2001-483447/52.	
XX	Human genome-derived single exon nucleic acid probes useful for analyzing	
PT	gene expression in human cervical epithelial cells.	
PS	Claim 27; SEQ ID NO 23367; 487bp; English.	
XX	The present invention relates to human single exon nucleic acid probes	
CC	(SENIP: see AAI10068-AAI28459). The present sequence is a peptide encoded	
CC	by one such probe. The SENIPs are derived from human HeLa cells. The SENIPs	
CC	can be used to produce a single exon microarray, which can be used for	
CC	measuring human gene expression in a sample derived from human cervical	
CC	epithelial cells. By measuring gene expression, the probes are therefore	
CC	useful in grading and/or staging of diseases of the cervix, notably	
CC	cervical cancer. Note: The sequence data for this patent did not form	
CC	part of the printed specification, but was obtained in electronic format	
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	
XX	Sequence 48 AA;	
XX	Query Match	100.0%; Score 19; DB 4; Length 48;
XX	Best Local Similarity	100.0%; Pred. No. 6.6e+02;
XX	Mismatches	4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 VAEF 4	
DB	34 VAEF 37	
DE		
DE	17-OCT-2001 (first entry)	
DE	Peptide #5033 encoded by probe for measuring placental gene expression.	
KW	Probe; microarray; human; placenta; antenatal diagnosis;	
KW	genetic disorder.	
OS	Homo sapiens.	

XX WO200157272-A2.  
 XX 09-AUG-2001.  
 XX 30-JAN-2001; 2001WO-US000663.  
 XX 04-FEB-2000; 2000US-0180312P.  
 XX 26-MAY-2000; 2000US-0207456P.  
 XX 30-JUN-2000; 2000US-00608408.  
 XX 03-AUG-2000; 2000US-00632366.  
 XX 21-SEP-2000; 2000US-0234687P.  
 XX 27-SEP-2000; 2000US-0236359P.  
 XX 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488897/53.  
 XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 XX gene expression in human placenta.  
 XX Claim 27; SEQ ID NO 31265; 654pp; English.  
 XX The present invention relates to single exon nucleic acid probes (SENP;  
 XX see AA131315-AA157546). The present sequence is a peptide encoded by one  
 XX such probe. The probes are useful for producing a microarray for  
 XX predicting, measuring and displaying gene expression in samples derived  
 XX from human placenta. The probes are useful for antenatal diagnosis of  
 XX human genetic disorders  
 XX Sequence 48 AA;  
 XX Query Match 100.0%; Score 19; DB 4; Length 48;  
 XX Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
 XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX 1 VAEF 4  
 XX |||||  
 XX 3 VAEF 6  
 XX RESULT 52  
 XX BB22868  
 XX D ABB22868 standard; protein; 48 AA.  
 XX C ABB22868;  
 XX T 23-JAN-2002 (first entry)  
 XX Protein #4867 encoded by probe for measuring heart cell gene expression.  
 XX Human; gene expression; heart; microarray; vascular system;  
 XX cardiovascular disease; hypertension; cardiac arrhythmia;  
 XX congenital heart disease.  
 XX Homo sapiens.  
 XX WO200157274-A2.  
 XX 09-AUG-2001.  
 XX 30-JAN-2001; 2001WO-US000666.  
 XX 04-FEB-2000; 2000US-0180312P.  
 XX 26-MAY-2000; 2000US-0207456P.  
 XX 30-JUN-2000; 2000US-00608408.  
 XX 03-AUG-2000; 2000US-00632366.  
 XX 21-SEP-2000; 2000US-0234687P.  
 XX 27-SEP-2000; 2000US-0236359P.  
 XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488899/53.  
 XX Single exon nucleic acid probes for analyzing gene expression in human  
 XX hearts.  
 XX Claim 15; SEQ ID NO 24639; 530pp; English.  
 XX The present invention relates to single exon nucleic acid probes for  
 XX measuring human gene expression in a sample derived from human heart (see  
 XX AB21535-AB41305). The present sequence is a protein encoded by one such  
 XX probe. The probes may be used for predicting, measuring and displaying  
 XX gene expression in samples derived from the human heart via microarrays.  
 XX By measuring gene expression, the probes are useful for predicting,  
 XX diagnosing, grading, staging, monitoring and prognosing diseases of the  
 XX human heart and vascular system e.g. cardiovascular disease,  
 XX hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
 XX sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WIPO  
 XX at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 48 AA;  
 XX Query Match 100.0%; Score 19; DB 4; Length 48;  
 XX Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
 XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX 1 VAEF 4  
 XX |||||  
 XX 3 VAEF 6  
 XX RESULT 53  
 XX AAM70687  
 XX ID AAM70687 standard; protein; 48 AA.  
 XX AC AAM70687;  
 XX DT 06-NOV-2001 (first entry)  
 XX Human bone marrow expressed probe encoded protein SEQ ID NO: 30993.  
 XX Human; bone marrow expressed exon; gene expression analysis; probe;  
 XX microarray; cancer; leukaemia; lymphoma; myeloma.  
 XX Homo sapiens.  
 XX WO200157276-A2.  
 XX 09-AUG-2001.  
 XX 30-JAN-2001; 2001WO-US000668.  
 XX 04-FEB-2000; 2000US-0180312P.  
 XX 26-MAY-2000; 2000US-0207456P.  
 XX 30-JUN-2000; 2000US-00608408.  
 XX 03-AUG-2000; 2000US-00632366.  
 XX 21-SEP-2000; 2000US-0234687P.  
 XX 27-SEP-2000; 2000US-0236359P.  
 XX 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488900/53.  
 XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 XX gene expression in human bone marrow.

XX Example 4; SEQ ID NO 30993; 658pp + Sequence Listing; English.  
PS  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention  
XX  
SQ Sequence 48 AA;  
Query Match 100.0%; Score 19; DB 4; Length 48;  
Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
DB 3 VAEF 6  
|||||  
RESULT 54  
ABG52389  
ID ABG52389 standard; peptide; 48 AA.  
XX AC  
XX ABG52389;  
XX  
DT 25-FEB-2003 (first entry)  
XX  
DE Human liver peptide, SEQ ID NO 31037.  
XX  
XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
KW hypercholesterolaemia; coronary heart disease.  
XX  
OS Homo sapiens.  
XX  
XX WO200157273-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 30-JAN-2001; 2001WO-US000664.  
PF  
XX 04-FEB-2000; 2000US-0180312P.  
PR  
XX 26-MAY-2000; 2000US-0207456P.  
PR  
XX 30-JUN-2000; 2000US-00608408.  
PR  
XX 03-AUG-2000; 2000US-00632366.  
PR  
XX 21-SEP-2000; 2000US-0234687P.  
PR  
XX 27-SEP-2000; 2000US-0236359P.  
PR  
XX 04-OCT-2000; 2000GB-00024263.  
PR  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-488898/53.  
XX  
DR Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human adult liver.  
XX  
XX Claim 27; SEQ ID NO 31037; 658pp; English.  
PS  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
XX measuring human gene expression in a sample derived from human adult  
XX liver, comprising one of 13109 defined nucleotide sequences given in the  
XX specification (or complements/ fragments). The probe hybridises at high  
XX stringency to a nucleic acid molecule expressed in the human adult liver.  
XX (I) may be used for predicting, measuring and displaying gene expression  
XX in samples derived from human adult liver. The genes identified may be  
XX involved in genetic liver diseases such as cirrhosis,  
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
XX associated with coronary heart disease. ABG47348-ABG59930 represent human  
XX liver single exon encoded peptides of the invention. Note: The sequence  
XX information for this patent does not appear in the printed specification

CC but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 48 AA;  
Query Match 100.0%; Score 19; DB 4; Length 48;  
Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
DB 3 VAEF 6  
|||||  
RESULT 55  
AAM06107  
ID AAM06107 standard; protein; 48 AA.  
XX AC  
XX AAM06107;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Peptide #4789 encoded by probe for measuring breast gene expression.  
XX  
XX Probe; human; breast disease; breast cancer; development disorder;  
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
XX Homo sapiens.  
XX  
XX WO200157270-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 29-JAN-2001; 2001WO-US000661.  
PF  
XX 04-FEB-2000; 2000US-0180312P.  
PR  
XX 26-MAY-2000; 2000US-0207456P.  
PR  
XX 30-JUN-2000; 2000US-00608408.  
PR  
XX 03-AUG-2000; 2000US-00632366.  
PR  
XX 21-SEP-2000; 2000US-0234687P.  
PR  
XX 27-SEP-2000; 2000US-0236359P.  
PR  
XX 04-OCT-2000; 2000GB-00024263.  
PR  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-476286/51.  
XX  
DR Novel single exon nucleic acid probe used to measuring gene expression in  
XX a human breast.  
XX  
XX Claim 27; SEQ ID NO 14847; 322pp; English.  
PS  
XX The present invention relates to novel single exon nucleic acid probes  
XX (see AAI00010-AAI10067). The present sequence is a peptide encoded by one  
XX such probe. The probes are useful for measuring human gene expression in  
XX a human breast sample, where the probe hybridises at high stringency to a  
XX nucleic acid expressed in the human breast. The probes are useful for  
XX predicting, diagnosing, grading, staging, monitoring and prognosing  
XX diseases of the human breast, particularly those diseases with polygenic  
XX aetiology. The diseases include: breast cancer, disorders of development,  
XX inflammatory diseases of the breast, fibrocystic changes, proliferative  
XX breast disease and non-carcinoma tumours. Note: The sequence data for  
XX this patent did not form part of the printed specification, but was  
XX obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 48 AA;  
Query Match 100.0%; Score 19; DB 4; Length 48;  
Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Y      1 VAEF 4
D      ||||
b      3 VAEF 6

RESULT 56
AOI0565
D      AAO10565 standard; protein; 53 AA.
C      AAO10565;
X
X
T      06-NOV-2001 (first entry)
E
X      Human polypeptide SEQ ID NO 24457.
W      Human; cytokine; cell proliferation; cell differentiation; gene therapy;
W      vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
W      tissue growth factor; immunomodulatory; cancer; leukaemia;
W      nervous system disorders; arthritis; inflammation.
X
S      Homo sapiens.
X
X      WO200164835-A2.
N
X
D      07-SEP-2001.
X
F      26-FEB-2001; 2001WO-US004927.
X
R      28-FEB-2000; 2000US-00515126.
R      18-MAY-2000; 2000US-00577409.
X
A      (HYSE-) HYSEQ INC.
X
X      Tang YT, Liu C, Drmanac RT;
I      WPI; 2001-514838/56.
X      N-PSDB; AAI90496.
X
T      Isolated nucleic acids and polypeptides, useful for preventing diagnosing
I      and treating e.g. leukemia, inflammation and immune disorders.
X
S      Claim 20; SEQ ID NO 24457; 1399pp + Sequence Listing; English.
X
X      The invention relates to human polynucleotides (AAI79941-AAI93841) and
X      the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
X      cytokine, cell proliferation or cell differentiation or which may induce
X      production of other cytokines in other cell populations. The
X      polynucleotides and polypeptides are useful in gene therapy, vaccines or
X      peptide therapy. The polypeptides have various cytokine-like activities,
X      e.g. stem cell growth factor activity, haematopoiesis regulating
X      activity, tissue growth factor activity, immunomodulatory activity and
X      activin/inhibin activity and may be useful in the diagnosis and/or
X      treatment of cancer, leukaemia, nervous system disorders, arthritis and
X      inflammation. Note: The sequence data for this patent did not form part
X      of the printed specification, but was obtained in electronic format
X      directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
X
Q      Sequence 53 AA;

Query Match      100.0%; Score 19; DB 4; Length 53;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      1 VAEF 4
b      33 VAEF 36

RESULT 57
AO05247
D      AAO05247 standard; protein; 53 AA.
X

```

XX PN WO200157182-A2.  
XX PD 09-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US001354.  
XX XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 18-AUG-2000; 2000US-0225759P.  
PR 22-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226682P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234323P.  
PR 21-SEP-2000; 2000US-0234324P.  
PR 25-SEP-2000; 2000US-0234397P.  
PR 23-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 05-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251988P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-02559678P.  
(HUMA-) HUMAN GENOME SCI INC.  
XX PA  
XX XX

I Rosen CA, Barash SC, Ruben SM;  
 X WPI; 2001-483426/52.  
 R N-PSDB; AAK60171.  
 X Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 T useful for preventing, diagnosing and/or treating cancers and metastasis.  
 X Claim 11; SEQ ID NO 14983; 3071pp + Sequence Listing; English.  
 X  
 S C AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 C amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 C activity, and can be used in gene therapy and vaccine production. (I)  
 C proteins and polynucleotides may be used in the prevention, diagnosis and  
 C treatment of diseases associated with inappropriate (I) expression. For  
 C example, they may be used to treat disorders associated with decreased  
 C expression by rectifying mutations or deletions in a patient's genome  
 C that affect the activity of (I) by expressing inactive proteins or to  
 C supplement the patient's own production of (I). Additionally, (I)  
 C polynucleotides may be used to produce the secreted (I), by inserting the  
 C nucleic acids into a host cell and culturing the cell to express the  
 C protein. (I) proteins and polynucleotides may be used to prevent,  
 C diagnose and treat immune/hematopoietic-related diseases, especially  
 C cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
 C to AAK87694 represent human immune/hematopoietic antigen genomic  
 C sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 C represent sequences used in the exemplification of the present invention.  
 X Q Sequence 56 AA;  
 Query Match 100.0%; Score 19; DB 4; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 1 VAEF 4  
 b 1 VAEF 4  
 E  
 D  
 X  
 C  
 X  
 C  
 X  
 T 24-JUN-2003 (first entry)  
 X Human matrix metalloproteinase 7 cleavage region peptide SEQ ID NO.4.  
 E  
 X Human; matrix metalloproteinase; MMP; anticancer; wound healing;  
 W matrix metalloproteinase inhibitor; antitumor; antiangiogenic; cardiant;  
 W vascular endothelial growth factor inhibitor; VEGF inhibitor; cytostatic;  
 W vulnary; cerebroprotective; antidiabetic; ophthalmological; tumour;  
 W dermatological; metastatic; non-metastatic; vascularised; heart disease;  
 W non-vascularised; surgical incision; chronic wound; stroke; angiogenesis;  
 W macular degeneration; diabetic retinopathy; cleavage region.  
 X Homo sapiens.  
 S WO2003018748-A2.  
 X  
 N  
 X  
 D 06-MAR-2003.  
 X  
 F 15-AUG-2002; 2002WO-US026319.  
 X  
 R 16-AUG-2001; 2001US-0312726P.  
 R 21-DEC-2001; 2001US-00032376.  
 R 21-MAY-2002; 2002US-00153185.  
 X  
 A (KIMB ) KIMBERLY-CLARK WORLDWIDE INC.  
 X Quirk S, Weart IF;  
 X

DR WPI; 2003-381408/36.  
 XX Anti-angiogenic composition comprising peptide inhibitor of matrix  
 PT metalloproteinase, useful for decreasing the expression of vascular  
 PT endothelial growth factor and treating cancers and tissue injuries.  
 XX Claim 17; Page 15; 103pp; English.  
 XX  
 CC The present invention describes an anti-angiogenic composition (I) for  
 CC inhibiting expression of vascular endothelial growth factor (VEGF). (I)  
 CC comprises an effective amount of a peptide inhibitor of matrix  
 CC metalloproteinase (MMP), where the peptide can inhibit the expression of  
 CC VEGF. (I) has cytostatic, vulnary, cardiant, cerebroprotective,  
 CC antidiabetic, ophthalmological and dermatological activities. (I) can be  
 CC used for inhibiting expression of VEGF, and so can be used for inhibiting  
 CC growth of tumours and diminishing tumours size. The tumour can be  
 CC metastatic, non-metastatic, vascularised, non-vascularised, hard or soft.  
 CC (I) is also useful for treating injuries including wounds, surgical  
 CC incisions, chronic wounds, heart diseases and stroke. (I) is also useful  
 CC for treating disorders characterised by excessive angiogenesis e.g.  
 CC macular degeneration and diabetic retinopathy. The present sequence  
 CC represents a human MMP cleavage region peptide, which is used in the  
 CC exemplification of the present invention  
 XX  
 SQ Sequence 56 AA;  
 Query Match 100.0%; Score 19; DB 6; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 VAEF 4  
 Db 31 VAEF 34  
 RESULT 60  
 ID ABG76312  
 XX ABG76312 standard; protein; 56 AA.  
 AC ABG76312;  
 DT 10-MAY-2003 (first entry)  
 XX Human matrix metalloproteinase (MMP) peptide inhibitor #4.  
 XX Human; peptide inhibitor; matrix metalloproteinase-7; MMP-7;  
 KW cleavage region; proenzyme form; cellular proliferation; fibroblast;  
 KW keratinocyte; healthy skin development; wound healing; scarring;  
 KW skin tone; wrinkle; anti-aging; vulnary.  
 XX Homo sapiens.  
 OS WO2003016520-A1.  
 XX  
 PN 27-FEB-2003.  
 PD  
 XX 15-AUG-2002; 2002WO-US026198.  
 PF  
 XX 16-AUG-2001; 2001US-0312726P.  
 PR 21-DEC-2001; 2001US-00032376.  
 PR 21-MAY-2002; 2002US-00153185.  
 XX (KIMB ) KIMBERLY-CLARK WORLDWIDE INC.  
 PA  
 XX Quirk S, Malik S, Villanueva JM;  
 FI WPI; 2003-289980/28.  
 XX  
 DR Novel peptide inhibitor of proteinase activity of matrix  
 PT metalloproteinases, e.g. matrix metalloproteinase-2, useful for  
 PT stimulating cellular proliferation of fibroblasts or keratinocytes.  
 XX Claim 1; Page 16; 120pp; English.  
 PS

XX The present invention relates to peptide inhibitors of metalloproteinases  
 CC (MMPs), particularly metalloproteinase-2 (MMP-2). The inhibitors have  
 CC peptide sequences related to the cleavage regions of the proenzyme forms  
 CC of the MMPs. The peptide inhibitors are useful for stimulating cellular  
 CC proliferation of fibroblasts or keratinocytes, promoting healthy skin  
 CC development, treating wounds, preventing scarring, improving skin tone,  
 CC reducing wrinkling and for simulating the development of smooth, healthy  
 CC skin. The peptide inhibitors are useful as anti-aging and wound healing  
 CC compounds. ABG76309-ABG76321 represent peptide inhibitors of MMPs  
 XX  
 SQ Sequence 56 AA;

Query Match 100.0%; Score 19; DB 6; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB 31 VAEF 34  
 |||||

RESULT 61  
 AAW55888  
 ID AAW55888 standard; protein; 58 AA.

XX AAW55888;

XX 22-JUL-1998 (first entry)

XX Rat PC12.

XX Rat; telomerase; human; cancer; screening; inhibitor; elucidation;  
 KW detection; probe; diagnosis; cell growth; ageing.

XX Rattus sp.

XX WO9807838-A1.

XX 26-FEB-1998.

XX 21-AUG-1997; 97WO-JP002904.

XX 21-AUG-1996; 96JP-00219761.

XX 31-JAN-1997; 97JP-00018878.

XX 17-FEB-1997; 97JP-00031807.

XX (MITU) MITSUBISHI CHEM CORP.

XX Ishikawa F, Nakamura H, Takahashi K, Fujino Y, Harada N;

XX WPI; 1998-169149/15.

XX N-PSDB; AAV25990.

XX Telomerase protein of higher animals and humans and gene encoding it -  
 PT for use in diagnosis of cancer, screening of telomerase inhibitors and  
 PT elucidation of biological control mechanisms.

XX Example 1; Page 69-70; 106pp; Japanese.

XX The present sequence represents rat PC12 which is used in an example of  
 CC the present invention which describes protein components of telomerase.  
 CC The DNA or RNA encoding the telomerase protein component or its fragments  
 CC can be used as a nucleotide probe for the detection of cancer cells and  
 CC for diagnosis of cancer. Potential telomerase inhibitors can be screened  
 CC by measuring their effect on the assay of the active form in cells or  
 CC tissues. The polypeptide and DNA coding for it can be used in the  
 CC elucidation of biological control mechanisms of, e.g. cell growth or  
 CC ageing and of the mechanisms of cancer development

XX Sequence 58 AA;

Query Match 100.0%; Score 19; DB 2; Length 58;

Best Local Similarity 100.0%; Pred. No. 8.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB 49 VAEF 52  
 |||||

RESULT 62  
 ABP04854  
 ID ABP04854 standard; protein; 58 AA.

XX ABP04854;

XX 24-JUN-2002 (first entry)

XX Human ORFX protein sequence SEQ ID NO:9690.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.

XX Homo sapiens.

XX WO200192523-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US010836.

XX 30-MAY-2000; 2000US-0206132P.

XX 29-AUG-2000; 2000US-0228716P.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach MD;

XX WPI; 2002-106308/14.

XX N-PSDB; ABN20606.

XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders.

XX Disclosure; SEQ ID NO 9690; 1037pp; English.

XX The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, diabetes mellitus, systemic  
 CC transplantation, cardiovascular diseases, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage. N.B. the sequence data for this patent did not



06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 27-AUG-1999; 99US-0151303P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 13-SEP-1999; 99US-0154018P.  
PR 15-SEP-1999; 99US-0154039P.  
PR 16-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157533P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158233P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159333P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 14-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161408P.  
PR 25-OCT-1999; 99US-0161409P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 19; DB 3; Length 59;  
Best Local Similarity 100.0%; Pred. No. 8.2e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
Db 45 VAEF 48  
RESULT 64  
AAG60683  
ID AAG60683 standard; protein; 59 AA.  
XX  
AC AAG60683;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 78631.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
06-SEP-2000.  
XX  
25-FEB-2000; 2000EP-00301439.  
XX  
25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137232P.  
PR 03-JUN-1999; 99US-0137548P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.

R 16-JUN-1999; 99US-0139453P.  
R 17-JUN-1999; 99US-0139492P.  
R 18-JUN-1999; 99US-0139454P.  
R 18-JUN-1999; 99US-0139455P.  
R 18-JUN-1999; 99US-0139456P.  
R 18-JUN-1999; 99US-0139457P.  
R 18-JUN-1999; 99US-0139458P.  
R 18-JUN-1999; 99US-0139459P.  
R 18-JUN-1999; 99US-0139460P.  
R 18-JUN-1999; 99US-0139461P.  
R 18-JUN-1999; 99US-0139462P.  
R 18-JUN-1999; 99US-0139463P.  
R 18-JUN-1999; 99US-0139750P.  
R 18-JUN-1999; 99US-0139763P.  
R 18-JUN-1999; 99US-0139817P.  
R 18-JUN-1999; 99US-0139899P.  
R 22-JUN-1999; 99US-0140353P.  
R 23-JUN-1999; 99US-0140354P.  
R 23-JUN-1999; 99US-0140695P.  
R 28-JUN-1999; 99US-0140823P.  
R 28-JUN-1999; 99US-0140891P.  
R 30-JUN-1999; 99US-0141287P.  
R 01-JUL-1999; 99US-0141842P.  
R 01-JUL-1999; 99US-0142154P.  
R 02-JUL-1999; 99US-0142055P.  
R 06-JUL-1999; 99US-0142390P.  
R 08-JUL-1999; 99US-0142803P.  
R 09-JUL-1999; 99US-0142820P.  
R 12-JUL-1999; 99US-0142977P.  
R 13-JUL-1999; 99US-0143542P.  
R 14-JUL-1999; 99US-0143624P.  
R 15-JUL-1999; 99US-0144005P.  
R 16-JUL-1999; 99US-0144085P.  
R 16-JUL-1999; 99US-0144086P.  
R 19-JUL-1999; 99US-0144325P.  
R 19-JUL-1999; 99US-0144331P.  
R 19-JUL-1999; 99US-0144332P.  
R 19-JUL-1999; 99US-0144333P.  
R 19-JUL-1999; 99US-0144334P.  
R 19-JUL-1999; 99US-0144335P.  
R 20-JUL-1999; 99US-0144352P.  
R 20-JUL-1999; 99US-0144352P.  
R 20-JUL-1999; 99US-0144864P.  
R 21-JUL-1999; 99US-0144814P.  
R 21-JUL-1999; 99US-0145086P.  
R 21-JUL-1999; 99US-0145088P.  
R 22-JUL-1999; 99US-0145085P.  
R 22-JUL-1999; 99US-0145087P.  
R 22-JUL-1999; 99US-0145089P.  
R 22-JUL-1999; 99US-0145192P.  
R 23-JUL-1999; 99US-0145145P.  
R 23-JUL-1999; 99US-0145218P.  
R 23-JUL-1999; 99US-0145224P.  
R 26-JUL-1999; 99US-0145276P.  
R 27-JUL-1999; 99US-0145913P.  
R 27-JUL-1999; 99US-0145918P.  
R 27-JUL-1999; 99US-0145919P.  
R 28-JUL-1999; 99US-0145951P.  
R 02-AUG-1999; 99US-0146386P.  
R 02-AUG-1999; 99US-0146388P.  
R 02-AUG-1999; 99US-0146389P.  
R 03-AUG-1999; 99US-0147038P.  
R 04-AUG-1999; 99US-0147204P.  
R 04-AUG-1999; 99US-0147302P.  
R 05-AUG-1999; 99US-0147192P.  
R 05-AUG-1999; 99US-0147260P.  
R 06-AUG-1999; 99US-0147303P.  
R 06-AUG-1999; 99US-0147303P.  
R 09-AUG-1999; 99US-0147433P.  
R 09-AUG-1999; 99US-0147935P.  
R 10-AUG-1999; 99US-0148171P.  
R 11-AUG-1999; 99US-0148319P.  
R 12-AUG-1999; 99US-0148341P.

PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 23-AUG-1999; 99US-0150566P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 18-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 22-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 19; DB 3; Length 59;  
Best Local Similarity 100.0%; Pred. No. 8.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VAEF 4  
45 VAEF 48

## RESULT 65

ABB42044  
ID ABB42044 standard; peptide; 60 AA.

XX AC ABB42044;  
XX DT 04-FEB-2002 (first entry)

XX DE Peptide #9550 encoded by human foetal liver single exon probe.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

XX PN WO200157277-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000669.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX PF WIPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing

XX PS Claim 27; SEQ ID NO 34679; 639pp + Sequence Listing; English.

XX CC The invention relates to a single exon nucleic acid probe for measuring  
XX CC human gene expression in a sample derived from human foetal liver. The  
XX CC single exon nucleic acid probes may be used for predicting, measuring and  
XX CC displaying gene expression in samples derived from human foetal liver. The  
XX CC present sequence is a peptide encoded by a single exon nucleic acid probe  
XX CC of the invention. Note: The sequence data for this patent did not form  
XX CC part of the printed specification, but was obtained in electronic format  
XX CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 60 AA;

Query Match 100.0%; Score 19; DB 4; Length 60;

Best Local Similarity 100.0%; Pred. No. 8.4e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4

DB 8 VAEF 11

## RESULT 66

AAW35846

ID AAW35846 standard; protein; 60 AA.

XX AC AAW35846;

XX DT 17-OCT-2001 (first entry)

XX DE Peptide #9893 encoded by probe for measuring placental gene expression.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

XX KW genetic disorder.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000663.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX PF WIPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing

XX PS Claim 27; SEQ ID NO 36115; 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENP:  
XX CC see AA31315-AA157546). The present sequence is a peptide encoded by one  
XX CC such probe. The probes are useful for producing a microarray for  
XX CC predicting, measuring and displaying gene expression in samples derived  
XX CC from human placenta. The probes are useful for antenatal diagnosis of  
XX CC human genetic disorders

XX SQ Sequence 60 AA;

Query Match 100.0%; Score 19; DB 4; Length 60;

Best Local Similarity 100.0%; Pred. No. 8.4e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4

DB 8 VAEF 11

## RESULT 67

AAW75737

ID AAW75737 standard; protein; 60 AA.

XX AC AAW75737;

XX DT 06-NOV-2001 (first entry)

XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 36043.

XX KW Human; bone marrow expressed exon; gene expression analysis; probe;

XX KW microarray; cancer; leukaemia; lymphoma; myeloma.

XX OS Homo sapiens.

XX PN WO200157276-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000668.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI; 2001-488900/53.  
Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human bone marrow.  
Example 4; SEQ ID NO 36043; 658pp + Sequence Listing; English.  
The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human bone marrow. They can be used to measure gene expression in bone marrow samples, which may enable the improved diagnosis and treatment of cancers such as lymphoma, leukaemia and myeloma. The present sequence is a protein encoded by one of the probes of the invention  
Sequence 60 AA;  
Query Match 100.0%; Score 19; DB 4; Length 60;  
Best Local Similarity 100.0%; Pred. NO. 8.4e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 VAEF 4  
|||||  
8 VAEF 11  
RESULT 68  
AM62925  
AA62925 standard; protein; 60 AA.  
AA62925;  
05-NOV-2001 (first entry)  
Human brain expressed single exon probe encoded protein SEQ ID NO: 35030.  
Human; brain expressed exon; gene expression analysis; probe; microarray;  
Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
Homo sapiens.  
WO200157275-A2.  
09-AUG-2001.  
30-JAN-2001; 2001WO-US000667.  
04-FEB-2000; 2000US-0180312P.  
26-MAY-2000; 2000US-0207456P.  
30-JUN-2000; 2000US-00608408.  
03-AUG-2000; 2000US-00632366.  
21-SEP-2000; 2000US-0234687P.  
27-SEP-2000; 2000US-0236359P.  
04-OCT-2000; 2000GB-00024263.  
(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI; 2001-483446/52.  
Single exon nucleic acid probes for analyzing gene expression in human brains.  
Example 4; SEQ ID NO 35030; 650pp + Sequence Listing; English.  
The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention  
XX Sequence 60 AA;  
SQ  
Query Match 100.0%; Score 19; DB 4; Length 60;  
Best Local Similarity 100.0%; Pred. NO. 8.4e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 VAEF 4  
|||||  
8 VAEF 11  
RESULT 69  
ABG57475  
ID ABG57475 standard; peptide; 60 AA.  
XX AC ABG57475;  
XX DT 25-FEB-2003 (first entry)  
XX DE Human liver peptide, SEQ ID NO 36123.  
XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
XX KW hypercholesterolaemia; coronary heart disease.  
XX OS Homo sapiens.  
XX PN WO200157273-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US000664.  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX PR 26-MAY-2000; 2000US-0207456P.  
XX PR 30-JUN-2000; 2000US-00608408.  
XX PR 03-AUG-2000; 2000US-00632366.  
XX PR 21-SEP-2000; 2000US-0234687P.  
XX PR 27-SEP-2000; 2000US-0236359P.  
XX PR 04-OCT-2000; 2000GB-00024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488998/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human adult liver.  
XX Claim 27; SEQ ID NO 36123; 658pp; English.  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for measuring human gene expression in a sample derived from human adult liver, comprising one of 13109 defined nucleotide sequences given in the specification (or complements/ fragments). The probe hybridises at high stringency to a nucleic acid molecule expressed in the human adult liver.  
XX (I) may be used for predicting, measuring and displaying gene expression in samples derived from human adult liver. The genes identified may be involved in genetic liver diseases such as cirrhosis, hypercholesterolaemia, hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is associated with coronary heart disease. ABG47348-ABG59930 represent human liver single exon encoded peptides of the invention. Note: The sequence information for this patent does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 60 AA;  
SQ  
Query Match 100.0%; Score 19; DB 4; Length 60;

Best Local Similarity 100.0%; Pred. No. 8.4e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 8 VAEF 11

RESULT 70  
AA17262  
ID AA17262 standard; peptide; 62 AA.  
XX AC AA17262;  
XX 09-AUG-1999 (first entry)  
DE HLH domain of drosophila E-spl m5.  
XX ORF polypeptide; dermatomyositis; polymyositis; inclusion body myositis;  
KW sarcoid myopathy; AZT myopathy; myocardial infarction; ischaemia; ERK-6;  
KW extracellular signal-regulated kinase; reperfusion.  
DS Drosophila melanogaster.  
XX WO9927099-A1.  
XX 03-JUN-1999.  
XX 23-NOV-1998; 98WO-US025046.  
XX 25-NOV-1997; 97US-0066848P.  
XX (SUGC-) SUGEN INC.  
PA Ullrich A, Giot J;  
PI WPI; 1999-357837/30.  
DR Nucleic acid encoding ORF polypeptide.  
XX Disclosure; Fig 2; 79pp; English.  
XX The invention relates to a human ORF polypeptide, a substrate for  
CC extracellular signal-regulated kinase, ERK-6. Substances that modulate  
CC the activity of the ORF polypeptide can be used to treat diseases  
CC selected from dermatomyositis, polymyositis, inclusion body myositis,  
CC sarcoid myopathy, AZT myopathy, myocardial infarction, and ischaemia/  
CC reperfusion. The probes and antibodies can be used to detect the presence  
CC of ORF in a sample  
XX Sequence 62 AA;  
XX Query Match 100.0%; Score 19; DB 2; Length 62;  
Best Local Similarity 100.0%; Pred. No. 8.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DY 1 VAEF 4  
DB 29 VAEF 32

RESULT 71  
AA17247  
ID AA17247 standard; protein; 62 AA.  
XX AC AA17247;  
XX 07-NOV-2001 (first entry)  
XX Human immune/haematopoietic antigen SEQ ID NO:18840.  
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis.

XX OS Homo sapiens.  
XX WO200157182-A2.  
XX 09-AUG-2001.  
XX 17-JAN-2001; 2001WO-US001354.  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 12-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 21-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.

R 26-SEP-2000; 2000US-0235484P.  
R 27-SEP-2000; 2000US-0235834P.  
R 27-SEP-2000; 2000US-0235836P.  
R 29-SEP-2000; 2000US-0236327P.  
R 29-SEP-2000; 2000US-0236367P.  
R 29-SEP-2000; 2000US-0236368P.  
R 29-SEP-2000; 2000US-0236369P.  
R 29-SEP-2000; 2000US-0236370P.  
R 29-SEP-2000; 2000US-0236802P.  
R 02-OCT-2000; 2000US-0237037P.  
R 02-OCT-2000; 2000US-0237038P.  
R 02-OCT-2000; 2000US-0237039P.  
R 02-OCT-2000; 2000US-0237040P.  
R 13-OCT-2000; 2000US-0239335P.  
R 13-OCT-2000; 2000US-0239337P.  
R 20-OCT-2000; 2000US-0240860P.  
R 20-OCT-2000; 2000US-0241221P.  
R 20-OCT-2000; 2000US-0241785P.  
R 20-OCT-2000; 2000US-0241786P.  
R 20-OCT-2000; 2000US-0241787P.  
R 20-OCT-2000; 2000US-0241808P.  
R 20-OCT-2000; 2000US-0241809P.  
R 20-OCT-2000; 2000US-0241826P.  
R 01-NOV-2000; 2000US-024617P.  
R 08-NOV-2000; 2000US-0246474P.  
R 08-NOV-2000; 2000US-0246475P.  
R 08-NOV-2000; 2000US-0246476P.  
R 08-NOV-2000; 2000US-0246477P.  
R 08-NOV-2000; 2000US-0246478P.  
R 08-NOV-2000; 2000US-0246523P.  
R 08-NOV-2000; 2000US-0246524P.  
R 08-NOV-2000; 2000US-0246525P.  
R 08-NOV-2000; 2000US-0246526P.  
R 08-NOV-2000; 2000US-0246527P.  
R 08-NOV-2000; 2000US-0246528P.  
R 08-NOV-2000; 2000US-0246532P.  
R 08-NOV-2000; 2000US-0246609P.  
R 08-NOV-2000; 2000US-0246610P.  
R 08-NOV-2000; 2000US-0246611P.  
R 08-NOV-2000; 2000US-0246613P.  
R 17-NOV-2000; 2000US-0249307P.  
R 17-NOV-2000; 2000US-0249308P.  
R 17-NOV-2000; 2000US-0249309P.  
R 17-NOV-2000; 2000US-0249310P.  
R 17-NOV-2000; 2000US-0249311P.  
R 17-NOV-2000; 2000US-0249312P.  
R 17-NOV-2000; 2000US-0249313P.  
R 17-NOV-2000; 2000US-0249314P.  
R 17-NOV-2000; 2000US-0249315P.  
R 17-NOV-2000; 2000US-0249316P.  
R 17-NOV-2000; 2000US-0249317P.  
R 17-NOV-2000; 2000US-0249318P.  
R 17-NOV-2000; 2000US-0249344P.  
R 17-NOV-2000; 2000US-0249345P.  
R 17-NOV-2000; 2000US-0249346P.  
R 17-NOV-2000; 2000US-0249347P.  
R 17-NOV-2000; 2000US-0249348P.  
R 17-NOV-2000; 2000US-0249349P.  
R 17-NOV-2000; 2000US-0249350P.  
R 01-DEC-2000; 2000US-0250160P.  
R 01-DEC-2000; 2000US-0250391P.  
R 05-DEC-2000; 2000US-0251030P.  
R 05-DEC-2000; 2000US-0251988P.  
R 05-DEC-2000; 2000US-0251989P.  
R 06-DEC-2000; 2000US-0251479P.  
R 08-DEC-2000; 2000US-0251856P.  
R 08-DEC-2000; 2000US-0251868P.  
R 08-DEC-2000; 2000US-0251869P.  
R 08-DEC-2000; 2000US-0251988P.  
R 08-DEC-2000; 2000US-0251990P.  
R 11-DEC-2000; 2000US-0254907P.  
R 05-JAN-2001; 2001US-0259678P.  
CX

PA (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
XX N-PSDB; AAK64028.  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX Claim 11; SEQ ID NO 18840; 3071pp + Sequence Listing; English.  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent, the  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX Sequence 62 AA;  
SQ  
  
Query Match 100.0%; Score 19; DB 4; Length 62;  
Best Local Similarity 100.0%; Pred. No. 8.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VAEF 4  
Db 25 VAEF 28  
  
RESULT 72  
ABM65667  
ID ABM65667 standard; protein; 62 AA.  
XX  
AC ABM65667;  
XX 20-OCT-2003 (first entry)  
XX  
XX Propionibacterium acnes immunogenic polypeptide #30343.  
DE  
XX Acne vulgaris; antisecborrheic; dermatological; antibacterial;  
KW immunostimulant; immune response; vaccine; immunogenic.  
XX  
XX Propionibacterium acnes.  
OS  
XX WO2003033515-A1.  
PN  
XX 24-APR-2003.  
PD  
XX 11-OCT-2002; 2002WO-US032727.  
PF  
XX 15-OCT-2001; 2001US-00978825.  
PR  
XX (CORI-) CORIXA CORP.  
XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
PI Barth B, Valliave-Douglas J;  
XX  
XX WPI; 2003-381789/36.  
DR  
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT

PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.

PS Claim 7; SEQ ID NO 30343; 1481pp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC polynucleotide of the invention; a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a specifically claimed P. acnes polypeptide which is  
 CC thought to contain an immunogenic region. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 62 AA;

Query Match 100.0%; Score 19; DB 6; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 8.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB 8 VAEF 11

RESULT 73  
 AAO09517  
 ID AAO09517 standard; protein; 64 AA.

XX AAO09517;  
 XX 06-NOV-2001 (first entry)  
 XX Human polypeptide SEQ ID NO 23409.

DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US0004927.

XX 28-FEB-2000; 2000US-00515126.

XX 18-MAY-2000; 2000US-00577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.  
 DR N-PSDB; AAI89448.

XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
 PT and treating e.g. leukemia, inflammation and immune disorders.

PS Claim 20; SEQ ID NO 23409; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 64 AA;

Query Match 100.0%; Score 19; DB 4; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB 54 VAEF 57

RESULT 74

AAI48352  
 ID AAI48352 standard; protein; 67 AA.

XX AAI48352;

XX 08-DEC-1999 (first entry)

XX Human prostate cancer-associated protein 49.

XX Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;  
 KW cancer; tissue specificity; human.

XX Homo sapiens.

XX DE19811194-A1.

XX 16-SEP-1999.

XX 10-MAR-1998; 98DE-01011194.

XX 10-MAR-1998; 98DE-01011194.

XX (META-) METAGEN GES GENOMFORSCHUNG MEH.

XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;

XX WPI; 1999-519629/44.

XX N-PSDB; AA233499.

XX New nucleic acid expressed at high level in normal prostatic tissue and  
 PT encoded polypeptides, used to treat cancer and screen for therapeutic  
 agents.

XX Claim 22; 143; 194pp; German.

XX This invention describes novel nucleic acid sequences (A) that are  
 CC expressed at high level in normal prostatic tissue. Polypeptides (I)  
 CC encoded by (A) are used: (a) for identifying agents for treatment of



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: May 24, 2004, 17:32:13 ; Search time 10 Seconds  
(without alignments)  
38.477 Million cell updates/sec

itle: US-09-594-978A-2

effect score: 19

equences: 1 VABP 4

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 283366 seqs, 96191526 residues

otal number of hits satisfying chosen parameters: 283366

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

atabase :

PIR 78:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	19	100.0	20	S39049	cytoxin-binding
2	19	100.0	33	G64625	hypothetical prote
3	19	100.0	48	E64618	hypothetical prote
4	19	100.0	52	A42375	hypothetical prote
5	19	100.0	54	E85745	unknown protein en
6	19	100.0	59	A69386	conserved hypotet
7	19	100.0	59	E71349	probable preprotei
8	19	100.0	63	JC4002	carboxymycin resista
9	19	100.0	63	G64007	hypothetical prote
10	19	100.0	64	I51350	transferrin - Atla
11	19	100.0	74	D97045	hypothetical prote
12	19	100.0	75	S07277	gene D protein - p
13	19	100.0	79	T18034	hypothetical prote
14	19	100.0	79	T27605	hypothetical prote
15	19	100.0	84	F83922	hypothetical prote
16	19	100.0	87	A38725	transferrin - shee
17	19	100.0	89	H64116	ribosomal protein
18	19	100.0	89	F86818	hypothetical prote
19	19	100.0	89	T15018	hypothetical prote
20	19	100.0	96	AB3053	hypothetical prote
21	19	100.0	97	AC1397	conserved hypotet
22	19	100.0	99	E83444	conserved hypotet
23	19	100.0	101	H90297	conserved hypotet
24	19	100.0	106	T10097	nifx protein - Met
25	19	100.0	108	C64490	hypothetical prote
26	19	100.0	109	E70949	probable regulator
27	19	100.0	112	T17582	hypothetical prote
28	19	100.0	112	B83431	type III export pr
29	19	100.0	114	A55872	kedarcidin (valida

30	19	100.0	114	2	D97414	hypothetical prote
31	19	100.0	115	2	D75218	hypothetical prote
32	19	100.0	116	1	IPAP	insulin precursor
33	19	100.0	118	2	T48682	hypothetical prote
34	19	100.0	119	2	F95327	hypothetical prote
35	19	100.0	122	2	F82861	conjugal transfer
36	19	100.0	125	2	B69129	gamma-carboxymuon
37	19	100.0	125	2	D70790	hypothetical prote
38	19	100.0	126	2	B81368	hypothetical prote
39	19	100.0	127	2	D70799	hypothetical prote
40	19	100.0	127	2	AB1732	hypothetical prote
41	19	100.0	128	2	A54797	ileal lipid-bindin
42	19	100.0	129	2	T02012	probable cathepsin
43	19	100.0	130	2	D72348	conserved hypotet
44	19	100.0	131	2	G75208	hypothetical prote
45	19	100.0	132	2	JQ2270	hypothetical 15.2K
46	19	100.0	132	2	AG1996	hypothetical prote
47	19	100.0	133	2	AG1996	conserved hypotet
48	19	100.0	133	2	AG1996	conserved hypotet
49	19	100.0	133	2	H56223	riboflavin synthas
50	19	100.0	134	1	WMEC15	ybgC protein - Esc
51	19	100.0	134	2	C90725	hypothetical prote
52	19	100.0	134	2	D85576	hypothetical prote
53	19	100.0	134	2	AD0592	conserved hypotet
54	19	100.0	136	2	S76416	hypothetical prote
55	19	100.0	137	2	E81355	hypothetical prote
56	19	100.0	138	2	F82106	flagellar basal-bo
57	19	100.0	139	2	A69052	riboflavin synthas
58	19	100.0	139	2	A83975	hypothetical prote
59	19	100.0	140	2	B69323	hypothetical prote
60	19	100.0	142	2	JQ2279	low-temperature re
61	19	100.0	142	2	JQ2280	low-temperature re
62	19	100.0	143	2	H55515	riboflavin synthas
63	19	100.0	143	2	T36978	probable transposa
64	19	100.0	145	1	B41715	ribosomal protein
65	19	100.0	145	2	S77043	hypothetical prote
66	19	100.0	146	2	F89837	conserved hypotet
67	19	100.0	147	2	C84294	hypothetical prote
68	19	100.0	147	2	T34803	hypothetical prote
69	19	100.0	148	2	E81788	conserved hypotet
70	19	100.0	148	2	S10655	hypothetical prote
71	19	100.0	149	2	AC1137	ribose 5-phosphate
72	19	100.0	150	2	AE2857	pseudouridin (impo
73	19	100.0	150	2	AB1089	protein gp35 from
74	19	100.0	150	2	AI1452	protein gp35 from
75	19	100.0	151	2	H87263	hypothetical prote
76	19	100.0	152	2	S04623	superoxide dismuta
77	19	100.0	154	2	E90184	hypothetical prote
78	19	100.0	155	2	I64130	ankyrin-repeat con
79	19	100.0	156	2	G81283	hypothetical prote
80	19	100.0	159	2	T17848	probable acetyltra
81	19	100.0	161	2	AE0981	hypothetical prote
82	19	100.0	163	2	A24504	salivary glue prot
83	19	100.0	163	2	G70524	hypothetical prote
84	19	100.0	164	2	AD1229	B. subtilis YabP
85	19	100.0	164	2	AE9293	hypothetical prote
86	19	100.0	166	2	S50509	hypothetical prote
87	19	100.0	168	2	S48866	hypothetical 19.5K
88	19	100.0	168	2	AE0192	probable C4-dicarb
89	19	100.0	169	2	T36387	probable acetyltra
90	19	100.0	169	2	AB2739	acetyltransferase
91	19	100.0	170	2	AC2917	invasion protein A
92	19	100.0	172	2	B75041	hypothetical prote
93	19	100.0	172	2	T49449	hemopoietic-specif
94	19	100.0	173	2	T22350	hypothetical prote
95	19	100.0	174	2	B70879	hypothetical prote
96	19	100.0	175	1	G71120	hypothetical prote
97	19	100.0	175	2	AC2648	conserved hypotet
98	19	100.0	175	2	I39055	Bcl-2 related - hu
99	19	100.0	176	2	C27805	ferritin chain M -
100	19	100.0	178	1	B69389	hypothetical prote
101	19	100.0	178	2	A38293	transcription fact
102	19	100.0	178	2	S03629	neurogenic gene co

103	19	100.0	179	2	S34345	hypothetical prote	176	19	100.0	219	2	H69199	conserved hypothet
104	19	100.0	180	2	D87031	probable acetyltra	177	19	100.0	219	2	AE1506	weakly carboxylest
105	19	100.0	180	2	C85358	hypothetical prote	178	19	100.0	220	2	T10379	protein tyrosine p
106	19	100.0	180	2	A10676	probable exported	179	19	100.0	220	2	B84381	riboflavin-specifi
107	19	100.0	181	2	AG2202	cobinamide kinase	180	19	100.0	221	2	C83947	hypothetical prote
108	19	100.0	183	2	S13186	plasma retinol-bin	181	19	100.0	223	2	D69311	conserved hypothet
109	19	100.0	184	2	A90903	probable ante-term	182	19	100.0	223	2	T35665	hypothetical prote
110	19	100.0	184	2	F90873	probable antitermi	183	19	100.0	223	2	AC0890	probable lipoprote
111	19	100.0	184	2	D71933	hypothetical prote	184	19	100.0	223	2	S69643	hypothetical prote
112	19	100.0	185	2	S43729	H+-transporting tw	185	19	100.0	224	2	AH1187	hypothetical prote
113	19	100.0	185	2	S93317	replication initia	186	19	100.0	226	2	AG3528	hypothetical prote
114	19	100.0	185	2	S93318	replication initia	187	19	100.0	228	2	A75390	chloramphenicol ac
115	19	100.0	185	2	D75635	resolvase - Deinoc	188	19	100.0	228	2	S48116	integral membrane
116	19	100.0	185	2	B97430	hypothetical prote	189	19	100.0	228	2	AD0693	pyridoxal kinase (
117	19	100.0	187	2	S16314	photosynthetic rea	190	19	100.0	228	2	D30857	hypothetical prote
118	19	100.0	187	2	AG2721	conserved hypothet	191	19	100.0	228	2	T10399	hypothetical prote
119	19	100.0	187	2	A58896	conserved hypothet	192	19	100.0	228	2	AH0081	probable exported
120	19	100.0	188	2	AB0132	conserved hypothet	193	19	100.0	229	2	F96721	probable glutathio
121	19	100.0	188	2	A71286	conserved hypothet	194	19	100.0	229	2	E70340	glutaredoxin-like
122	19	100.0	188	2	H85713	unknown protein en	195	19	100.0	230	2	E97440	DNA-3-methyladenin
123	19	100.0	189	2	T29159	hypothetical prote	196	19	100.0	230	2	S73229	ribosomal protein
124	19	100.0	189	2	H97633	pseudouracil precu	197	19	100.0	231	2	AF1144	phosphoglycerate m
125	19	100.0	190	2	A29413	ubiquinol-cytochro	198	19	100.0	231	2	AF1503	phosphoglycerate m
126	19	100.0	191	2	AG2392	transcription regu	199	19	100.0	231	2	C83179	conserved hypothet
127	19	100.0	192	2	D97503	hypothetical prote	200	19	100.0	231	2	AF3380	hypothetical prote
128	19	100.0	193	1	C64328	conserved hypothet	201	19	100.0	232	2	E70553	hypothetical prote
129	19	100.0	194	1	RPCR5	resolvase - Escher	202	19	100.0	232	2	E87520	hypothetical prote
130	19	100.0	194	2	C84033	hypothetical prote	203	19	100.0	234	2	F96620	hypothetical prote
131	19	100.0	194	2	F97691	hypothetical prote	204	19	100.0	234	2	AC2185	hypothetical prote
132	19	100.0	194	2	B87453	RNA polymerase sig	205	19	100.0	234	2	E91119	hypothetical prote
133	19	100.0	195	2	B62004	hypothetical prote	206	19	100.0	234	2	D85964	hypothetical prote
134	19	100.0	196	1	BVECAU	transcription regu	207	19	100.0	234	2	S22360	hypothetical prote
135	19	100.0	196	2	AH1853	hexose phosphate t	208	19	100.0	235	2	E97753	(p)ppGpp 3-pyropho
136	19	100.0	196	2	AH0953	Two-component syst	209	19	100.0	235	2	C95848	probable transcrip
137	19	100.0	196	2	F91204	transcription regu	210	19	100.0	236	2	A98351	agropine synthesis
138	19	100.0	196	2	H86050	hypothetical prote	211	19	100.0	236	2	AE2931	agropine synthesis
139	19	100.0	196	2	A64888	probable resolvase	212	19	100.0	236	2	E70728	hypothetical prote
140	19	100.0	196	2	D64909	probable resolvase	213	19	100.0	236	2	A84686	hypothetical prote
141	19	100.0	196	2	T21347	hypothetical prote	214	19	100.0	236	2	T50908	hypothetical prote
142	19	100.0	197	2	C86748	hypothetical prote	215	19	100.0	237	2	F83699	two-component resp
143	19	100.0	198	2	S48749	cytochrome P460 pr	216	19	100.0	237	2	C97085	hypothetical prote
144	19	100.0	199	1	VAHU	plasma retinol-bin	217	19	100.0	238	2	C86656	glycerol uptake fa
145	19	100.0	199	2	H90206	amidotransferase h	218	19	100.0	239	2	D81218	conserved hypothet
146	19	100.0	200	2	AB0620	probable bacteriop	219	19	100.0	239	2	F81795	probable periplasm
147	19	100.0	201	1	VARB	retinol-binding pr	220	19	100.0	239	2	H75195	hypothetical prote
148	19	100.0	201	2	A39486	plasma retinol-bin	221	19	100.0	239	2	S31023	gene 88 protein -
149	19	100.0	201	2	I46257	retinol binding pr	222	19	100.0	240	2	D69261	sulfate ABC transp
150	19	100.0	201	2	T00799	hypothetical prote	223	19	100.0	240	2	B84606	probable Alp synth
151	19	100.0	202	2	T35126	hypothetical prote	224	19	100.0	241	2	T14808	hypothetical prote
152	19	100.0	202	2	S77556	hypothetical prote	225	19	100.0	242	2	B70366	hypothetical prote
153	19	100.0	202	2	AH1169	hypothetical prote	226	19	100.0	242	2	G97959	H+-transporting tw
154	19	100.0	202	2	C97866	hypothetical prote	227	19	100.0	242	2	S43728	probable beta-keto
155	19	100.0	204	2	G89753	protein p11C7.1 [i	228	19	100.0	248	2	AF0177	water channel prot
156	19	100.0	205	2	S30739	hypothetical prote	229	19	100.0	248	2	T12632	probable tonoplast
157	19	100.0	205	2	H91225	hypothetical prote	230	19	100.0	248	2	T01648	probable high-affi
158	19	100.0	205	2	F85072	hypothetical prote	231	19	100.0	248	2	F72604	hypothetical prote
159	19	100.0	205	2	G01942	mitotic feedback c	232	19	100.0	248	2	T18315	serine/threonine p
160	19	100.0	208	2	AH2658	3-methyladenine-DN	233	19	100.0	249	2	G81693	conserved hypothet
161	19	100.0	208	2	AG6642	hypothetical prote	234	19	100.0	249	2	B69343	hypothetical prote
162	19	100.0	209	2	H97519	conserved hypothet	235	19	100.0	249	2	T19088	hypothetical prote
163	19	100.0	212	2	AB3521	conserved hypothet	236	19	100.0	250	2	T01012	TobRR7-18C protein
164	19	100.0	213	2	G82067	hypothetical prote	237	19	100.0	250	2	S13719	probable membrane
165	19	100.0	213	2	AF2551	transcription regu	238	19	100.0	250	2	S51781	hypothetical prote
166	19	100.0	213	2	AB4248	probable GTP bindi	239	19	100.0	250	2	D64580	(R)-hydroxyglutary
167	19	100.0	214	2	T30310	transcription regu	240	19	100.0	251	2	F69494	hypothetical prote
168	19	100.0	215	2	F84889	hypothetical prote	241	19	100.0	252	2	C90422	ABC transporter, p
169	19	100.0	218	2	H83485	conserved hypothet	242	19	100.0	252	2	F87575	hypothetical prote
170	19	100.0	218	2	T03287	osmotin protein ho	243	19	100.0	253	1	S17711	probable dehydroge
171	19	100.0	218	2	D71392	coat protein - cuc	244	19	100.0	253	2	C65067	2-deoxy-D-gluconat
172	19	100.0	218	2	AB0254	probable fumarylac	245	19	100.0	253	2	C91091	2-deoxy-D-gluconat
173	19	100.0	218	2	T47889	hypothetical prote	246	19	100.0	253	2	F85936	2-deoxy-D-gluconat
174	19	100.0	218	2	T47889	hypothetical prote	247	19	100.0	253	2	AD0210	2-deoxy-D-gluconat
175	19	100.0	219	2	T35128	hypothetical prote	248	19	100.0	253	2		

249	19	100.0	253	2	B73552	probable ATP-depen	322	19	100.0	291	2	G69479	methionyl aminopep
250	19	100.0	254	2	F90087	ATP-dependent Clp	323	19	100.0	291	2	E71491	probable geranyl t
251	19	100.0	255	2	T02078	photosystem II oxy	324	19	100.0	291	2	H96971	N-terminal Chey re
252	19	100.0	256	2	AE0161	deoxyribose operon	325	19	100.0	292	1	G64503	conserved hypothet
253	19	100.0	257	2	S12985	phosphoprotein pho	326	19	100.0	292	2	T32000	hypothetical prote
254	19	100.0	258	2	T52130	probable copper/zinc	327	19	100.0	293	2	F71344	hypothetical prote
255	19	100.0	259	2	T35273	probable enoyl coA	328	19	100.0	294	2	H64465	methionyl aminopep
256	19	100.0	260	2	S56108	26S proteasome reg	329	19	100.0	294	2	S48934	nucleic acid-bindi
257	19	100.0	261	2	B83275	conserved hypothet	330	19	100.0	295	2	D82341	ribosomal protein
258	19	100.0	262	2	F69479	nicotinate-nucleot	331	19	100.0	295	2	A49906	glucose-1-phosphat
259	19	100.0	263	2	T36359	hypothetical prote	332	19	100.0	295	2	C82828	glucose-1-phosphat
260	19	100.0	264	2	S75599	hypothetical prote	333	19	100.0	295	2	C43718	hypothetical prote
261	19	100.0	265	2	G81427	periplasmic protei	334	19	100.0	297	2	F64470	sulfate permease (
262	19	100.0	266	2	AF2054	hypothetical prote	335	19	100.0	297	2	AC2002	transposase alr156
263	19	100.0	267	2	AD2895	short-chain dehydr	336	19	100.0	297	2	AD1931	transposase alr099
264	19	100.0	268	2	G97670	probable oxidoredu	337	19	100.0	297	2	AD2141	transposase alr268
265	19	100.0	269	2	T00100	hypothetical prote	338	19	100.0	297	2	AG2142	transposase alr266
266	19	100.0	270	2	G84010	hypothetical prote	339	19	100.0	297	2	AD2138	transposase alr338
267	19	100.0	271	2	H98213	transcription repr	340	19	100.0	297	2	A12228	transposase alr362
268	19	100.0	272	2	B72333	bacteriocin - Ther	341	19	100.0	297	2	A12258	transposase alr362
269	19	100.0	273	2	AC0132	probable aldo/keto	342	19	100.0	298	2	H98170	myo-inositol catab
270	19	100.0	274	2	A57490	matrilysin (BC 3.4	343	19	100.0	298	2	AE3116	transcription regu
271	19	100.0	275	2	T24397	hypothetical prote	344	19	100.0	299	2	G82393	NAD(P)-arginine AD
272	19	100.0	276	2	A87404	ribosomal protein	345	19	100.0	300	2	JC4367	osteopontin - rat
273	19	100.0	277	2	A13121	ABC transporter, m	346	19	100.0	301	2	JC5811	protein T23E18.2 (
274	19	100.0	278	2	F98165	hypothetical prote	347	19	100.0	302	2	A96789	cytochrome-c perox
275	19	100.0	279	1	B5BY13	major intrinsic pr	348	19	100.0	304	2	G81417	hypothetical prote
276	19	100.0	280	2	B44499	erythrocyte integr	349	19	100.0	305	2	G40927	hypothetical prote
277	19	100.0	281	2	A41616	water channel prot	350	19	100.0	305	2	G71441	hypothetical prote
278	19	100.0	282	2	JC1320	uterine water chan	351	19	100.0	306	2	B98164	glycosyltransferas
279	19	100.0	283	2	I52366	hypothetical prote	352	19	100.0	306	2	AE3123	hypothetical prote
280	19	100.0	284	2	C71228	signal recognition	353	19	100.0	307	2	TL9582	thioredoxin-disulf
281	19	100.0	285	2	A56487	water channel prot	354	19	100.0	308	2	D6742	dnaj protein homol
282	19	100.0	286	2	JC2348	conserved hypothet	355	19	100.0	309	1	S62835	conserved hypothet
283	19	100.0	287	2	B89937	amvC protein - The	356	19	100.0	309	2	D90520	hypothetical prote
284	19	100.0	288	2	S37705	protein F4H10.8 (	357	19	100.0	309	2	S76393	glutaminyl-peptid
285	19	100.0	289	2	H86690	hypothetical prote	358	19	100.0	309	2	T09564	phosphoprotein pho
286	19	100.0	290	2	C83709	acetyl-CoA carboxy	359	19	100.0	310	2	T03596	conserved hypothet
287	19	100.0	291	2	H97337	dihydroneopterin a	360	19	100.0	310	2	C90245	conserved hypothet
288	19	100.0	292	2	B97260	uroporphyrinogen-I	361	19	100.0	311	2	A95117	dihydrodipicolinat
289	19	100.0	293	2	AG1842	hypothetical prote	362	19	100.0	311	2	G97986	dihydrodipicolinat
290	19	100.0	294	2	T19845	azarene carbazole	363	19	100.0	311	2	A93196	purine nucleosidas
291	19	100.0	295	2	JW0102	conserved hypothet	364	19	100.0	311	2	B84341	hypothetical prote
292	19	100.0	296	2	A72420	hypothetical prote	365	19	100.0	312	2	S24264	phosphoprotein pho
293	19	100.0	297	2	B81983	conserved hypothet	366	19	100.0	312	2	S31086	phosphoprotein pho
294	19	100.0	298	2	AS1040	conserved hypothet	367	19	100.0	312	2	S31089	cysteine proteinas
295	19	100.0	299	2	E71293	probable nicotinam	368	19	100.0	312	2	B23705	hypothetical prote
296	19	100.0	300	2	S52582	prephenate dehydra	369	19	100.0	312	2	S67052	dnak-type molecula
297	19	100.0	301	2	F86842	prephenate dehydra	370	19	100.0	313	2	A26485	phosphoenolpyruvat
298	19	100.0	302	2	T22051	hypothetical prote	371	19	100.0	313	2	S27698	desiccation-relate
299	19	100.0	303	2	D84185	hypothetical prote	372	19	100.0	313	2	E45509	hypothetical prote
300	19	100.0	304	2	S71506	site-specific DNA-	373	19	100.0	313	2	E95853	glycerate dehydrog
301	19	100.0	305	2	AC3579	transcription regu	374	19	100.0	313	2	AD1285	glycerate dehydrog
302	19	100.0	306	2	C86221	hypothetical prote	375	19	100.0	314	2	AG1656	phosphoprotein pho
303	19	100.0	307	2	B84637	fof1-type ATP synt	376	19	100.0	314	2	S52371	phosphoprotein pho
304	19	100.0	308	2	F97252	tRNA-pseudouridine	377	19	100.0	314	2	H69458	potassium channel
305	19	100.0	309	2	B70200	hypothetical prote	378	19	100.0	314	2	AG2517	hypothetical prote
306	19	100.0	310	2	H83551	2-hydroxyacetate-	379	19	100.0	316	2	S26225	phosphoprotein pho
307	19	100.0	311	2	T31275	probable aldose re	380	19	100.0	316	2	S42397	DNA-(apurinic or a
308	19	100.0	312	2	T37996	hypothetical prote	381	19	100.0	316	2	T32993	hypothetical prote
309	19	100.0	313	2	B70703	envelope-like prot	382	19	100.0	317	1	AT25917	osteopontin precu
310	19	100.0	314	2	AC4494	2-hydroxyacetate-	383	19	100.0	317	2	T03594	phosphoprotein pho
311	19	100.0	315	2	JC5419	pyridoxal kinase (	384	19	100.0	317	2	AC39500	DNA-(apurinic or a
312	19	100.0	316	2	AI0288	glutamine-tRNA lig	385	19	100.0	317	2	F82672	conserved hypothet
313	19	100.0	317	2	S07563	probable pyridoxal	386	19	100.0	317	2	D69258	phosphoprotein pho
314	19	100.0	318	2	F64520	pyridoxal kinase 2	387	19	100.0	318	2	S20882	phosphoprotein pho
315	19	100.0	319	2	A90922	pyridoxal kinase 2	388	19	100.0	318	2	TL3015	DNA-(apurinic or a
316	19	100.0	320	2	B85770	hypothetical prote	389	19	100.0	318	2	S23550	ubiquitin-activat
317	19	100.0	321	2	T35229	hypothetical prote	390	19	100.0	319	2	T05515	protein T12C24.6 (
318	19	100.0	322	2	AI3400	glycine hydroxymet	391	19	100.0	320	2	D86259	unknown protein P2
319	19	100.0	323	2	G64167	hypothetical prote	392	19	100.0	320	2	D96750	probable oxidoredu
320	19	100.0	324	2	AD2870	ABC transporter, m	393	19	100.0	320	2	G96834	
321	19	100.0	325	2	F97646	probable aliphatic	394	19	100.0	320	2		

395	19	100.0	320	2	T28992	hypotheical prote	468	19	100.0	352	2	AC1560	two-component sens
396	19	100.0	322	2	D8785	hypotheical prote	469	19	100.0	352	2	AE1202	two-component sens
397	19	100.0	323	2	T09550	phosphoprotein pho	470	19	100.0	353	2	D97295	parvulin-like pept
398	19	100.0	323	2	G84325	hypotheical prote	471	19	100.0	355	2	AG1877	hypotheical prote
399	19	100.0	324	2	E84505	hypotheical prote	472	19	100.0	355	2	AI3624	spemidine/putresc
400	19	100.0	325	2	JC7553	brain mitochondria	473	19	100.0	357	2	AC1276	aminopeptidase hom
401	19	100.0	326	2	G69366	homoserine dehydro	474	19	100.0	357	2	AC1639	aminopeptidase hom
402	19	100.0	326	2	T05094	peroxidase homolog	475	19	100.0	357	2	AE4551	hypotheical prote
403	19	100.0	326	2	T19070	hypotheical prote	476	19	100.0	359	2	T02011	probable cathepsin
404	19	100.0	327	2	AC3482	sulfate-binding pr	477	19	100.0	359	2	D95316	probable ABC-trans
405	19	100.0	327	2	E95968	hypotheical expor	478	19	100.0	359	2	D95382	probable ABC-trans
406	19	100.0	328	2	E75564	ABC transporter, A	479	19	100.0	360	2	D69025	pleiotropic regula
407	19	100.0	328	2	A89867	hypotheical prote	480	19	100.0	360	2	AB3078	hypotheical prote
408	19	100.0	328	2	C70368	conserved hypotet	481	19	100.0	360	2	G98208	probable ATP-bind
409	19	100.0	328	2	H84548	hypotheical prote	482	19	100.0	361	2	AB3042	hypotheical prote
410	19	100.0	329	2	AD2629	33 kDa chaperonin,	483	19	100.0	361	2	D98243	opuaa (AF234615) [
411	19	100.0	329	2	T50117	mo25 homolog limpo	484	19	100.0	361	2	AE1979	ABC transporter Ar
412	19	100.0	329	2	T29218	hypotheical prote	485	19	100.0	361	2	E96743	probable deoxyuan
413	19	100.0	330	1	H64404	hypotheical prote	486	19	100.0	362	2	C97448	hypotheical prote
414	19	100.0	330	2	AB3417	probable oxidoredu	487	19	100.0	363	2	B95371	probable oxidoredu
415	19	100.0	331	2	T35499	probable D-lactate	488	19	100.0	363	2	AE1134	acyl-CoA dehydrog
416	19	100.0	331	2	AE2928	dehydrogenase Atu3	489	19	100.0	363	2	C81888	probable acyl-CoA
417	19	100.0	331	2	D82972	probable lipolytic	490	19	100.0	364	2	G70364	conserved hypotet
418	19	100.0	331	2	AE2666	divalent cation tr	491	19	100.0	365	2	S76544	hypotheical prote
419	19	100.0	333	2	D75067	probable dehydrog	492	19	100.0	366	1	G49964	photosynthetic rea
420	19	100.0	333	2	C71165	probable dehydrog	493	19	100.0	366	2	T50891	cytochrome subunit
421	19	100.0	333	2	AI1881	collagenase PrtC (	494	19	100.0	366	2	G84249	NADH-dependent fla
422	19	100.0	334	1	S15318	transcription regu	495	19	100.0	368	2	G70462	succinate-CoA liga
423	19	100.0	334	2	T49195	hypotheical prote	496	19	100.0	368	2	G97291	hypotheical prote
424	19	100.0	334	2	AE5951	hypotheical prote	497	19	100.0	368	2	S75852	ABC-type transport
425	19	100.0	335	2	D28807	HK203J protein -	498	19	100.0	368	2	C83898	response regulator
426	19	100.0	336	2	B95906	probable dehydrog	499	19	100.0	368	2	A97189	lps biosynthesis p
427	19	100.0	336	2	AE0337	probable aspartate	500	19	100.0	369	2	S56638	mitogen-activated
428	19	100.0	336	2	AB3569	ABC transporter Ar	501	19	100.0	369	2	A95124	rNA polymerase sig
429	19	100.0	336	2	C87623	cobalamin biosynth	502	19	100.0	369	2	C97994	rNA polymerase sig
430	19	100.0	336	2	T00571	dolichyl-phosphate	503	19	100.0	369	2	B83571	probable ATP-bind
431	19	100.0	337	2	T35862	probable secreted	504	19	100.0	369	2	S77028	protein kinase, 41
432	19	100.0	337	2	T22757	hypotheical prote	505	19	100.0	369	2	T51477	glutamine-rich pro
433	19	100.0	338	2	AG2938	2-hydroxyacid-fam	506	19	100.0	370	2	C36819	Cl4 protein - rabb
434	19	100.0	338	2	H98343	hypotheical prote	507	19	100.0	370	2	F95927	probable mandelate
435	19	100.0	338	2	E75301	conserved hypotet	508	19	100.0	370	2	JC7998	platelet-derived g
436	19	100.0	339	2	S17930	transcription init	509	19	100.0	371	2	T09622	protein kinase MMK
437	19	100.0	339	2	JC7509	glycoprotein Vi-1	510	19	100.0	371	2	F86625	GTP-binding protei
438	19	100.0	339	2	T35630	hypotheical prote	511	19	100.0	371	2	D95000	GTP-binding protei
439	19	100.0	340	2	JC1125	phosphotransferase	512	19	100.0	371	2	D97872	conserved hypotet
440	19	100.0	340	2	T32646	hypotheical prote	513	19	100.0	371	2	AE0686	hypotheical prote
441	19	100.0	340	2	E83146	membrane-bound lyt	514	19	100.0	372	2	AD0075	probable transport
442	19	100.0	340	2	I49451	alpha 4 protein -	515	19	100.0	373	2	AC0253	ribonuclease iii (
443	19	100.0	341	2	S71223	xyloglucan endo-1,	516	19	100.0	375	2	AB3101	dipeptidase [impor
444	19	100.0	342	2	AE0087	sigma-54 transcrip	517	19	100.0	375	2	H98185	membrane dipeptida
445	19	100.0	342	2	S29894	strictosidine synt	518	19	100.0	376	2	T16116	hypotheical prote
446	19	100.0	342	2	AD3258	cobS protein limpo	519	19	100.0	377	2	D95861	genome polypeptide
447	19	100.0	343	2	B86446	probable endoxylg	520	19	100.0	377	2	D95861	probable ABC trans
448	19	100.0	343	2	C88986	protein C50H11.7 [	521	19	100.0	378	1	B38178	tellA protein - pla
449	19	100.0	343	2	D84390	sulfate transport	522	19	100.0	379	2	AF2409	mannosyl transfera
450	19	100.0	344	1	RWETC2	T-cell surface gly	523	19	100.0	379	2	H82284	queuine tRNA-ribos
451	19	100.0	344	2	AE8354	hypotheical prote	524	19	100.0	379	2	G97201	NAD(PAD)-depend
452	19	100.0	344	2	B8967	T-cell surface gly	525	19	100.0	379	2	JC7710	NADH-rubredoxin ox
453	19	100.0	344	2	I49585	CD2 antigen protei	526	19	100.0	379	2	AD3350	phnM protein limpo
454	19	100.0	344	2	AI1902	4-hydroxyphenylpyr	527	19	100.0	379	2	AE3225	ABC transporter, m
455	19	100.0	344	2	S01325	strictosidine synt	528	19	100.0	380	2	B64533	cystathionine gamm
456	19	100.0	345	2	F90239	anthranilate phosp	529	19	100.0	380	2	D71973	probable cystathio
457	19	100.0	345	2	E83927	hypotheical prote	530	19	100.0	381	2	S46583	442K curved dna-bi
458	19	100.0	346	2	AE4588	MHC class I histoc	531	19	100.0	381	2	AG3304	pleiotropic regula
459	19	100.0	346	2	C95404	hypotheical prote	532	19	100.0	383	2	G72777	probable S2P metal
460	19	100.0	346	2	C90590	hypotheical prote	533	19	100.0	384	2	D86821	hydroxymethylgluta
461	19	100.0	347	2	D69373	immunogenic protei	534	19	100.0	385	2	T16447	hypotheical prote
462	19	100.0	347	2	A26645	flagellar motor sw	535	19	100.0	385	2	H84411	hypotheical prote
463	19	100.0	347	2	F97427	flagellar motor sw	536	19	100.0	386	2	D64074	phosphoglycerate k
464	19	100.0	348	2	E69162	sulfate transport	537	19	100.0	386	2	B97411	33K chaperonin (he
465	19	100.0	349	2	H72714	probable O-sialogl	538	19	100.0	387	2	AC3190	hypotheical prote
466	19	100.0	351	2	D84541	hypotheical prote	539	19	100.0	388	2	E70894	probable metB prot
467	19	100.0	352	2	S22464	strictosidine synt	540	19	100.0	388	2	F87208	cystathionine [gam

541	19	100.0	388	2	T23670	hypothetical prote	614	19	100.0	427	2	G98259	methionine gamma-1
542	19	100.0	389	2	T15102	hypothetical prote	615	19	100.0	428	2	F69050	glycine hydroxymet
543	19	100.0	389	2	G84245	NADH dehydrogenase	616	19	100.0	429	2	G72323	phosphorylruvate hy
544	19	100.0	390	2	E92272	conserved hypotnet	617	19	100.0	429	2	S30049	transcription fact
545	19	100.0	390	2	H55023	nicotinate phospho	618	19	100.0	429	2	AC2227	hypothetical prote
546	19	100.0	390	2	A84226	farnesyl-diphospha	619	19	100.0	430	2	C83975	O-acetylhomoserine
547	19	100.0	390	2	T25996	hypothetical prote	620	19	100.0	430	2	B95892	probable ABC trans
548	19	100.0	390	2	D86291	hypothetical prote	621	19	100.0	431	2	C69087	dihydroliipoamide d
549	19	100.0	391	2	C70972	probable trehalose	622	19	100.0	431	2	G96987	enolase [imported]
550	19	100.0	392	2	B84017	probable RNA methy	623	19	100.0	431	2	T21594	hypothetical prote
551	19	100.0	392	2	E25880	telomere-binding p	624	19	100.0	431	2	D70185	3-hydroxy-3-methyl
552	19	100.0	393	2	S95261	serine proteinase	625	19	100.0	431	2	S37775	filamin, muscle -
553	19	100.0	393	2	C89961	hypothetical prote	626	19	100.0	432	1	S15203	glycine hydroxymet
554	19	100.0	394	2	B70411	fimbrial assembly	627	19	100.0	432	2	S08277	cyclin A - human
555	19	100.0	394	2	C81333	probable efflux pu	628	19	100.0	432	2	AB2403	ATP-binding protei
556	19	100.0	394	2	S76331	hypothetical prote	629	19	100.0	433	1	B70677	cytochrome P450 Rv
557	19	100.0	395	2	S38812	cyclin A - chicken	630	19	100.0	433	2	B70465	probable GTP bindi
558	19	100.0	395	2	T20724	hypothetical prote	631	19	100.0	434	2	S30334	glycine hydroxymet
559	19	100.0	397	2	H84225	acyl-CoA dehydroge	632	19	100.0	434	2	D71480	probable flagellum
560	19	100.0	397	2	B98127	serine proteinase	633	19	100.0	436	2	G69466	transcription fact
561	19	100.0	397	2	A72315	hypothetical prote	634	19	100.0	437	1	A31752	tetrahydrofolylpol
562	19	100.0	398	2	C95201	hydroxymethylgluta	635	19	100.0	437	2	C64113	high-affinity gluc
563	19	100.0	398	2	A98068	hypothetical prote	636	19	100.0	437	2	JC4988	transcription fact
564	19	100.0	399	2	A84972	nicotinate phospho	637	19	100.0	437	2	S42111	transcription fact
565	19	100.0	399	2	H87342	Glycosyl transfera	638	19	100.0	438	2	F95333	probable aminotran
566	19	100.0	399	2	F83796	Multidrug-efflux t	639	19	100.0	438	2	A98161	hypothetical prote
567	19	100.0	399	2	AG2221	hypothetical prote	640	19	100.0	438	2	H86006	hypothetical prote
568	19	100.0	400	2	T03460	probable leucine/i	641	19	100.0	438	2	AB0997	high-affinity gluc
569	19	100.0	402	1	A70707	cytochrome P450-Rv	642	19	100.0	439	2	AC0462	glycerol-3-phospha
570	19	100.0	402	2	H82813	ABC transporter, A	643	19	100.0	439	2	AC2948	hypothetical prote
571	19	100.0	403	2	H82440	NADH dehydrogenase	644	19	100.0	439	2	G98334	glycerol-3-phospha
572	19	100.0	405	2	G84222	probable metallope	645	19	100.0	440	2	AC2985	conserved hypotnet
573	19	100.0	405	2	A83084	cyclin A - bovine	646	19	100.0	440	2	D98238	probable sugar iso
574	19	100.0	406	2	S24788	serine proteinase	647	19	100.0	441	2	C87233	probable cell inva
575	19	100.0	406	2	F84152	tryptophan oxygena	648	19	100.0	442	2	T44655	O-acetylhomoserine
576	19	100.0	406	2	G20222	hypothetical prote	649	19	100.0	443	2	A99657	hypothetical prote
577	19	100.0	406	2	T30748	hypothetical prote	650	19	100.0	443	2	A85508	hypothetical prote
578	19	100.0	407	2	B71679	protein kinase, 48	651	19	100.0	443	2	T32088	hypothetical prote
579	19	100.0	408	2	T08069	hypothetical prote	652	19	100.0	444	2	C70444	biotin carboxylase
580	19	100.0	408	2	B97741	hypothetical prote	653	19	100.0	445	2	T16025	hypothetical prote
581	19	100.0	409	2	AD0492	Multidrug transloc	654	19	100.0	445	2	S32036	finger protein XFG
582	19	100.0	409	2	T02776	y4dM protein - Rhi	655	19	100.0	445	2	B40970	undulin 2 - human
583	19	100.0	409	2	F86151	P22M8.10 protein -	656	19	100.0	446	2	C82140	C4-dicarboxylate t
584	19	100.0	410	2	B48585	transcription fact	657	19	100.0	447	2	C96497	glyceraldhyde-3-p
585	19	100.0	410	2	B48585	transcription fact	658	19	100.0	448	2	AP0437	conserved hypotnet
586	19	100.0	411	2	H83355	hypothetical prote	659	19	100.0	449	2	D70726	probable gabT - My
587	19	100.0	411	2	AS0199	peptidase T (EC 3.	660	19	100.0	449	2	C70846	probable merC prot
588	19	100.0	412	1	RNECTA	tRNA adenylitrans	661	19	100.0	449	2	AP2646	chemotaxis motD pr
589	19	100.0	412	1	RNECTA	ferredoxin reducta	662	19	100.0	449	2	E97428	glucoamylase motD pr
590	19	100.0	412	2	C81121	tRNA nucleotidyl t	663	19	100.0	450	2	T39433	permease VCA0904 l
591	19	100.0	412	2	B55966	hypothetical prote	664	19	100.0	450	2	F82402	hypothetical prote
592	19	100.0	412	2	T21419	phosphoenolpyruvac	665	19	100.0	450	2	T21593	hypothetical prote
593	19	100.0	413	2	C69160	transcription fact	666	19	100.0	450	2	C86466	hypothetical prote
594	19	100.0	413	2	T48364	Glutamyl-tRNA redu	667	19	100.0	453	2	S37935	hypothetical prote
595	19	100.0	415	2	A81920	Glutamyl-tRNA redu	668	19	100.0	453	2	C83008	conserved hypotnet
596	19	100.0	415	2	C81183	cyclin A2 - Africa	669	19	100.0	455	2	B71335	probable purine-bi
597	19	100.0	415	2	I51637	diaminopimelate de	670	19	100.0	456	2	T46986	hypothetical prote
598	19	100.0	416	2	D95231	diaminopimelate de	671	19	100.0	456	2	AH0240	probable chlorohyd
599	19	100.0	416	2	F98095	hypothetical prote	672	19	100.0	458	2	S75328	DNA-directed DNA p
600	19	100.0	418	2	H72203	hypothetical prote	673	19	100.0	459	2	S33424	repSA protein - St
601	19	100.0	418	2	T35993	probable aminotran	674	19	100.0	460	2	A12218	hypothetical prote
602	19	100.0	419	2	B85035	hypothetical prote	675	19	100.0	461	1	WZRTKS	argininosuccinate
603	19	100.0	420	2	S65084	finger protein XFG	676	19	100.0	461	2	E71672	fumarate hydratase
604	19	100.0	421	2	T47393	hypothetical prote	677	19	100.0	462	2	A33313	malonyl-CoA decarb
605	19	100.0	421	2	T47393	hypothetical prote	678	19	100.0	462	2	I51699	gene XGF 5.1c prot
606	19	100.0	422	2	C83063	Glutamyl-tRNA redu	679	19	100.0	463	2	D97826	fumarate hydratase
607	19	100.0	422	2	S38501	cyclin A2 - mouse	680	19	100.0	464	1	WZHU8S	argininosuccinate
608	19	100.0	422	2	S37280	cyclin A - mouse	681	19	100.0	464	1	WZHU8S	probable sulfatase
609	19	100.0	423	2	T19145	hypothetical prote	682	19	100.0	465	2	F70837	replication initia
610	19	100.0	424	2	T13506	citrate (si)-synth	683	19	100.0	466	2	T36212	nicotinamide nucle
611	19	100.0	424	2	H87520	hypothetical prote	684	19	100.0	467	2	A12231	hypothetical prote
612	19	100.0	426	2	H75407	transcription term	685	19	100.0	467	2	G82697	hypothetical prote
613	19	100.0	427	2	AB3025	methionine gamma-1	686	19	100.0	468	2	S39832	probable phosphopr
										469	1	KCPGI	interstitial colla

687	19	100.0	469	2	S74825	probable Rieske ir	760	19	100.0	513	2	A85697	Na+/H+ antiporter,
688	19	100.0	470	2	S50083	phosphatase - short	761	19	100.0	513	2	G64864	Na+/H+-exchanging
689	19	100.0	471	2	S27676	mannose-1-phosphat	762	19	100.0	514	2	S46733	hypothetical prote
690	19	100.0	472	2	C82917	preprotein translo	763	19	100.0	514	2	T10559	hypothetical prote
691	19	100.0	473	2	T04738	hypothetical prote	764	19	100.0	514	2	AG0723	regulator of intra
692	19	100.0	474	2	E47677	guanosine diphosph	765	19	100.0	517	2	T29852	hypothetical prote
693	19	100.0	475	2	S22621	phosphonamutase	766	19	100.0	517	2	T27101	hypothetical prote
694	19	100.0	476	2	D69503	conserved hypothet	767	19	100.0	518	1	A27705	alpha-amylase (EC
695	19	100.0	477	2	T24197	hypothetical prote	768	19	100.0	518	2	T46887	presol methylhyd
696	19	100.0	478	2	S75530	hydrogenase large	769	19	100.0	522	2	A31556	glucose transport
697	19	100.0	479	2	AG3210	proline dipeptidas	770	19	100.0	522	2	B34067	Na+/H+-exchanging
698	19	100.0	480	2	B49651	long-chain-fatty-a	771	19	100.0	523	2	S06920	glucose transport
699	19	100.0	481	2	G87292	hypothetical prote	772	19	100.0	524	1	T40127	IMP dehydrogenase
700	19	100.0	482	2	C64119	starch synthase (S	773	19	100.0	524	2	A31318	glucose transporte
701	19	100.0	483	2	C95177	hypothetical prote	774	19	100.0	524	2	T06134	hypothetical prote
702	19	100.0	484	2	S71360	noea protein - Rhi	775	19	100.0	524	2	T08931	hypothetical prote
703	19	100.0	485	2	H95313	Noea host specific	776	19	100.0	528	2	S5205	prochlorophyllid
704	19	100.0	486	2	JC4386	adenylyl cyclase-a	777	19	100.0	528	2	T31459	hypothetical prote
705	19	100.0	487	2	J38409	adenylyl cyclase-a	778	19	100.0	528	2	JC4814	Na+/H+-exchanging
706	19	100.0	488	2	T23792	hypothetical prote	779	19	100.0	529	2	AF3059	cytochrome d oxida
707	19	100.0	489	2	S15312	rfbm protein - Sal	780	19	100.0	529	2	B98227	tubulin delta chai
708	19	100.0	490	2	AH0725	mannose-1-phosphat	781	19	100.0	532	2	T07903	deoxyribodipyrimid
709	19	100.0	491	2	B86367	protein F28f24.16	782	19	100.0	532	2	S52046	methyl-accepting c
710	19	100.0	492	2	H96729	probable alanine a	783	19	100.0	536	2	C82433	4-coumarate-CoA li
711	19	100.0	493	2	A70190	hypothetical prote	784	19	100.0	537	2	T09755	4-coumarate-CoA li
712	19	100.0	494	2	T10095	nife protein - Met	785	19	100.0	537	2	T09710	2,3-dihydroxybenzo
713	19	100.0	495	2	T47422	cellulase-like pro	786	19	100.0	537	2	AF0575	conserved hypothet
714	19	100.0	496	2	D82446	probable formate t	787	19	100.0	537	2	H87502	IMP cyclohydrolyase
715	19	100.0	497	2	T36302	probable 2-dehydro	788	19	100.0	538	2	AD3281	bifunctional purin
716	19	100.0	498	2	E86416	unknown protein, 3	789	19	100.0	538	2	AF2323	hypothetical prote
717	19	100.0	499	2	S95238	PTS system, membra	790	19	100.0	540	2	S76584	vibriobactin-speci
718	19	100.0	500	2	H83131	5-carboxy-2-hydrox	791	19	100.0	540	2	H82282	probable chemotaxi
719	19	100.0	501	2	C96018	probable starch sy	792	19	100.0	545	2	AB0691	fumarate hydratase
720	19	100.0	502	2	T49017	hypothetical prote	793	19	100.0	548	2	AB1854	glucose-6-phosphat
721	19	100.0	503	2	T39456	zinc finger protei	794	19	100.0	548	2	C81089	glucose-6-phosphat
722	19	100.0	504	2	A35926	probable argininos	795	19	100.0	548	2	T23270	acetyl choline rec
723	19	100.0	505	2	T09734	1-aminocyclopropan	796	19	100.0	548	2	A90253	hypothetical prote
724	19	100.0	506	2	T23055	hypothetical prote	797	19	100.0	549	2	T24508	hypothetical prote
725	19	100.0	507	2	H90423	sugar transport pr	798	19	100.0	549	2	JC4083	glucose-6-phosphat
726	19	100.0	508	2	T44576	L-2,4-diaminobutyr	799	19	100.0	549	2	B82330	glucose-6-phosphat
727	19	100.0	509	2	E71486	probable s/t prote	800	19	100.0	550	2	T24019	hypothetical prote
728	19	100.0	510	2	S37033	glucose-6-phosphat	801	19	100.0	550	2	T22557	hypothetical prote
729	19	100.0	511	2	T11894	NADH2 dehydrogenas	802	19	100.0	551	2	B84106	hypothetical prote
730	19	100.0	512	2	B64642	cell division protc	803	19	100.0	551	2	S03828	regulatory protein
731	19	100.0	513	2	T01235	hypothetical prote	804	19	100.0	553	2	AG3604	cytochrome d ubiqu
732	19	100.0	514	2	C71873	septum formation p	805	19	100.0	553	2	S1HQMA	malate synthase (E
733	19	100.0	515	2	H70573	hypothetical prote	806	19	100.0	555	1	C87472	steroid monooxygen
734	19	100.0	516	2	S10134	plasmid recombinat	807	19	100.0	556	2	T16790	hypothetical prote
735	19	100.0	517	2	A33982	58k mobilization p	808	19	100.0	556	2	T29503	serine C-palmitoyl
736	19	100.0	518	2	T27693	hypothetical prote	809	19	100.0	558	2	A75216	hypothetical prote
737	19	100.0	519	2	A22275	hypothetical prote	810	19	100.0	558	2	T28881	hypothetical prote
738	19	100.0	520	2	S27785	acetyl-CoA acetyl	811	19	100.0	558	2	A86813	xylan 1,4-beta-xy
739	19	100.0	521	2	AG3134	N-ethylamine ch	812	19	100.0	559	2	AB0984	probable membrane
740	19	100.0	522	2	T14236	NADH2 dehydrogenas	813	19	100.0	559	2	H82532	conserved hypothet
741	19	100.0	523	2	F88359	protein Y5H1A.5 (	814	19	100.0	559	2	G81842	30S ribosomal prot
742	19	100.0	524	2	E98153	N-ethylamine ch	815	19	100.0	561	2	C81099	30S ribosomal prot
743	19	100.0	525	2	B40669	fructose phosphotr	816	19	100.0	561	2	H69086	ABC transporter -
744	19	100.0	526	2	H86146	hypothetical prote	817	19	100.0	561	2	H97497	hypothetical prote
745	19	100.0	527	2	C91251	probable tail shea	818	19	100.0	562	2	B82351	methyl-accepting c
746	19	100.0	528	2	B18163	exopolysphatase	819	19	100.0	564	2	B72860	probable type II D
747	19	100.0	529	2	G81917	exopolysphatase	820	19	100.0	565	2	S69779	adhesin Ape5-2 pre
748	19	100.0	530	2	F18107	probable ubiquinol	821	19	100.0	567	2	T16105	hypothetical prote
749	19	100.0	531	2	B81184	ubiquinone biosynt	822	19	100.0	567	2	D72359	conserved hypothet
750	19	100.0	532	2	A82193	Sun/nucleolar prot	823	19	100.0	567	2	C90485	beta-glucuronidase
751	19	100.0	533	2	S27113	malonyl-CoA decarb	824	19	100.0	570	2	T06737	hypothetical prote
752	19	100.0	534	2	A84244	probable signaling	825	19	100.0	571	2	S33853	hydrogenase (EC 1
753	19	100.0	535	2	T16903	hypothetical prote	826	19	100.0	576	1	B71959	proline-tRNA ligas
754	19	100.0	536	2	C70468	phosphoribosylam	827	19	100.0	576	2	C69059	acetolactate synth
755	19	100.0	537	2	T31831	hypothetical prote	828	19	100.0	577	2	F64549	proline-tRNA ligas
756	19	100.0	538	2	A00475	probable magnesium	829	19	100.0	577	2	T30583	probable peptide s
757	19	100.0	539	2	G98102	hypothetical prote	830	19	100.0	580	2	C86530	31 ribosomal prote
758	19	100.0	540	2	T14638	cytochrome P450 Cy	831	19	100.0	580	2	D72093	ribosomal protein
759	19	100.0	541	2	A99839	Na+/H+ antiporter	832	19	100.0	580	2		

833	19	100.0	582	2	T16104	hypothetical prote	906	19	100.0	639	2	A56126	peroxisomal target
834	19	100.0	583	2	G86481	Mukator-like trans	907	19	100.0	640	1	A55073	transforming prote
835	19	100.0	583	2	T46177	villin 3 homolog T	908	19	100.0	640	2	F75553	hypothetical prote
836	19	100.0	587	2	E82431	methy1-accepting c	909	19	100.0	641	2	AD3502	dnak protein limpo
837	19	100.0	588	2	T25248	hypothetical prote	910	19	100.0	642	2	S55521	beta-fructofuranos
838	19	100.0	588	2	A43740	DG42 protein - Afr	911	19	100.0	644	2	JQ0160	3-isopropylmalate
839	19	100.0	588	2	A33460	hypothetical cytos	912	19	100.0	644	2	A45635	dnak-type molecula
840	19	100.0	589	2	A46459	macrophage-activat	913	19	100.0	645	2	JC5642	dnak-type molecula
841	19	100.0	589	2	S74568	hypothetical prote	914	19	100.0	645	2	T25824	hypothetical prote
842	19	100.0	590	2	B81104	nitrate/nitrite se	915	19	100.0	646	2	B82295	hypothetical prote
843	19	100.0	590	2	C81911	nitrate/nitrite se	916	19	100.0	647	2	AF1488	transcription anti
844	19	100.0	591	2	S43506	hypothetical prote	917	19	100.0	648	2	AQ0314	probable thiamin p
845	19	100.0	591	2	S70524	guanine nucleotide	918	19	100.0	649	2	JN0809	drebrin E (Clone 9
846	19	100.0	592	2	S64323	probable membrane	919	19	100.0	650	2	F72540	hypothetical prote
847	19	100.0	592	2	T45827	pectinesterase-lik	920	19	100.0	651	2	T12083	beta-fructofuranos
848	19	100.0	598	2	B71095	hypothetical prote	921	19	100.0	651	2	JC4610	dnak-type molecula
849	19	100.0	599	2	T32430	hypothetical prote	922	19	100.0	651	2	T14763	hypothetical prote
850	19	100.0	602	2	D69331	probable DNA topoi	923	19	100.0	652	2	I39811	parasporeal crystal
851	19	100.0	602	2	T01360	hypothetical prote	924	19	100.0	654	2	S70905	transferrin-bindin
852	19	100.0	603	2	T39674	hypothetical dnaJ	925	19	100.0	654	2	AD3183	peptidase limpothe
853	19	100.0	604	2	F87936	protein M01G12.12	926	19	100.0	656	1	B70107	hypothetical prote
854	19	100.0	604	2	T23669	hypothetical prote	927	19	100.0	660	2	S70904	transferrin-bindin
855	19	100.0	605	2	I39837	dnak-type molecula	928	19	100.0	669	2	C82984	ATP-dependent DNA
856	19	100.0	605	2	D84687	hypothetical prote	929	19	100.0	669	2	C96630	hypothetical prote
857	19	100.0	605	2	S46833	hypothetical prote	930	19	100.0	670	2	D88457	hypothetical prote
858	19	100.0	606	2	T35710	transferase - Stre	931	19	100.0	670	2	AH2425	hypothetical prote
859	19	100.0	607	2	B86744	Dnak protein limpo	932	19	100.0	672	2	S61157	hypothetical prote
860	19	100.0	607	2	S39342	dnak-type molecula	933	19	100.0	673	2	C83080	probable chemotaxi
861	19	100.0	607	2	B95060	dnak protein limpo	934	19	100.0	674	2	E84300	bacterio-opsin act
862	19	100.0	607	2	G97928	hypothetical prote	935	19	100.0	674	2	T22733	hypothetical prote
863	19	100.0	607	2	A95898	probable sensor ki	936	19	100.0	678	1	RDRT04	NADPH-ferrihemopro
864	19	100.0	610	2	S41315	hypothetical prote	937	19	100.0	678	2	B48089	growth arrest-spec
865	19	100.0	610	2	C89939	Dnak protein limpo	938	19	100.0	680	2	A42297	peptidyl-dipeptida
866	19	100.0	610	2	G69130	conserved hypothet	939	19	100.0	680	2	AH0678	dipeptidyl carboxy
867	19	100.0	612	2	A70756	probable ABC-type	940	19	100.0	684	2	T25603	hypothetical prote
868	19	100.0	613	2	T43738	dnak-type molecula	941	19	100.0	685	2	S64985	hypothetical prote
869	19	100.0	613	2	AE1621	class I heat-shock	942	19	100.0	686	2	C64428	hypothetical prote
870	19	100.0	613	2	A11258	class I heat-shock	943	19	100.0	688	2	E86409	hypothetical prote
871	19	100.0	614	2	S77221	protein kinase pkn	944	19	100.0	688	2	D96930	transferrin - Atia
872	19	100.0	619	2	A84411	hypothetical prote	945	19	100.0	690	2	T11749	transferrin - Atia
873	19	100.0	621	2	T06717	phosphothricin-t	946	19	100.0	691	2	T32748	hypothetical prote
874	19	100.0	622	2	PT0059	two-component sens	947	19	100.0	691	2	AD2766	conserved hypothet
875	19	100.0	622	2	E83160	parasporeal crystal	948	19	100.0	691	2	H97546	hypothetical 81.9K
876	19	100.0	622	2	S17402	hypothetical prote	949	19	100.0	694	1	TFRBP	transferrin precu
877	19	100.0	624	2	C83389	hypothetical prote	950	19	100.0	696	1	S01384	transferrin - pig
878	19	100.0	625	2	D64107	asparagine synthas	951	19	100.0	696	2	E96831	hypothetical prote
879	19	100.0	625	2	D86903	probable periplasm	952	19	100.0	696	2	H83024	hypothetical prote
880	19	100.0	625	2	G96976	carnitine O-acetyl	953	19	100.0	698	1	TFHUP	transferrin precu
881	19	100.0	627	2	S53369	hypothetical prote	954	19	100.0	703	2	G82249	asma protein VC103
882	19	100.0	627	2	S68958	hypothetical prote	955	19	100.0	704	2	I47228	carbonic anhydrase
883	19	100.0	627	2	D71374	carbon-monoxide de	956	19	100.0	705	2	JC5873	phenylalanine ammo
884	19	100.0	628	1	H64390	hypothetical prote	957	19	100.0	706	2	D86227	hypothetical prote
885	19	100.0	628	2	B64534	nucleoside-triphos	958	19	100.0	706	2	S33761	transferrin precu
886	19	100.0	631	1	NPV217	nucleoside-triphos	959	19	100.0	707	2	S29029	phenylalanine ammo
887	19	100.0	631	2	H36847	nucleoside-triphos	960	19	100.0	707	2	S60588	phenylalanine ammo
888	19	100.0	631	2	T28539	hypothetical prote	961	19	100.0	708	2	T14295	conserved hypothet
889	19	100.0	631	2	C72163	Old protein - vari	962	19	100.0	708	2	JC5872	phenylalanine ammo
890	19	100.0	631	2	T37384	nucleoside triphos	963	19	100.0	708	2	AD2195	ATP-dependent DNA
891	19	100.0	631	2	S70908	transferrin-bindin	964	19	100.0	708	2	JC5872	phenylalanine ammo
892	19	100.0	634	2	T37072	hypothetical prote	965	19	100.0	710	2	E83226	phenylalanine ammo
893	19	100.0	634	2	S32349	probable SNF2-type	966	19	100.0	712	2	T03663	ATP-dependent DNA
894	19	100.0	634	2	B64112	dnak-type molecula	967	19	100.0	712	2	T01858	phenylalanine ammo
895	19	100.0	635	2	B82273	dnak protein VC085	968	19	100.0	712	2	A41464	phenylalanine ammo
896	19	100.0	635	2	AP1881	1-deoxy-xylulose 5	969	19	100.0	714	2	A41464	phenylalanine ammo
897	19	100.0	636	1	TVMSMB	transforming prote	970	19	100.0	714	2	T35770	hypothetical prote
898	19	100.0	636	1	AA7042	dnak-type molecula	971	19	100.0	715	1	JX0199	hemolysin secretio
899	19	100.0	637	2	B83052	Dnak repair helicas	972	19	100.0	715	2	GH3269	fatty-acid beta-ox
900	19	100.0	637	2	E75044	dnak-type molecula	973	19	100.0	715	4	TVMSMY	transforming prote
901	19	100.0	638	1	IOECDK	dnak-type molecula	974	19	100.0	716	2	AH2319	glycyl-tRNA synthet
902	19	100.0	638	2	F90630	heat shock protein	975	19	100.0	716	2	D69855	conserved hypothet
903	19	100.0	638	2	F85481	dnak-type molecula	976	19	100.0	716	2	G83627	hypothetical prote
904	19	100.0	638	2	AE0503	Dnak protein (heat	977	19	100.0	721	2	A70764	probable glgX prot
905	19	100.0	639	2	S72163	methy1-accepting c	978	19	100.0	722	2	S76367	hypothetical prote

979 19 100.0 724 2 B85045 probable calcium c  
 980 19 100.0 725 1 ZBPA4 rapid lysis protei  
 981 19 100.0 725 2 S52990 phenylalanine ammo  
 982 19 100.0 725 2 G84787 phenylalanine ammo  
 983 19 100.0 726 2 A10147 probable ATP-depen  
 984 19 100.0 729 2 A91038 hypothetical prote  
 985 19 100.0 729 2 AF0808 probable membrane  
 986 19 100.0 729 2 AF0808 dipeptidyl peptida  
 987 19 100.0 738 2 A87515 hypothetical prote  
 988 19 100.0 742 2 D85882 NADH2 dehydrogenas  
 989 19 100.0 744 2 T13048 genome polyprotein  
 990 19 100.0 746 2 S31816 hypothetical prote  
 991 19 100.0 748 2 G96498 hypothetical prote  
 992 19 100.0 754 2 T25551 hypothetical prote  
 993 19 100.0 756 2 T04187 subtilisin-like pr  
 994 19 100.0 761 1 TVHUVB transforming prote  
 995 19 100.0 763 2 S23457 polysulfide reduct  
 996 19 100.0 764 2 S49849 aconitate hydratase  
 997 19 100.0 768 2 H82367 sensory box/GSDEF  
 998 19 100.0 768 2 H65013 yfeA protein - Esc  
 999 19 100.0 772 2 T05838 subtilisin-like pr  
 1000 19 100.0 774 2 G71308 probable aminopept  
 1000 19 100.0 781 2 A69312 DNA polymerase B1

## ALIGNMENTS

RESULT 1  
 S39049  
 Cytoxin-binding protein - rabbit  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C>Date: 18-Feb-1994 #sequence\_revision 19-Apr-1996 #text\_change 16-Feb-1997  
 C/Accession: S39049  
 R:Jantz, F.; Mohr, M.; Grimmig, M.; Leidolf, R.; Linder, D.  
 Eur. J. Biochem. 217, 1123-1128, 1993  
 A>Title: Pseudomonas aeruginosa cytoxin-binding protein in rabbit erythrocyte membrane  
 A/Reference number: S39049; MUID:94039134; PMID:7693466  
 A/Accession: S39049  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 1-20 <IUT>  
 C:Superfamily: lens fiber membrane major intrinsic protein

Query Match 100.0%; Score 19; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 14 VAEF 17

RESULT 2  
 364625  
 Hypothetical protein HP0847 - Helicobacter pylori (strain 26695)  
 C:Species: Helicobacter pylori #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
 C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
 C/Accession: G64625  
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997  
 A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A>Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A/Reference number: A64520; MUID:97394467; PMID:9252185  
 A/Accession: G64625  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-33 <TOM>  
 A/Cross-references: GB:A5000596; GB:A5000511; NID:g2313982; PIDN:AAD07920.1; PID:g231400

Query Match 100.0%; Score 19; DB 2; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 88;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 ||||  
 DB 15 VAEF 18  
 RESULT 3  
 E64618  
 Hypothetical protein HP0789 - Helicobacter pylori (strain 26695)  
 C:Species: Helicobacter pylori  
 C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
 C/Accession: E64618  
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997  
 A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A>Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A/Reference number: A64520; MUID:97394467; PMID:9252185  
 A/Accession: E64618  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-48 <TOM>  
 A/Cross-references: GB:A5000591; GB:A5000511; NID:g2313918; PIDN:AAD07847.1; PID:g231393

Query Match 100.0%; Score 19; DB 2; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 15 VAEF 18

RESULT 4  
 A42375  
 Hypothetical protein (lon 5' region) - Bacillus brevis  
 C:Species: Bacillus brevis  
 C>Date: 10-Jul-1992 #sequence\_revision 10-Jul-1992 #text\_change 15-Oct-1999  
 C/Accession: A42375; I39873  
 R:Ito, K.; Ueda, S.; Yamagata, H.  
 J. Bacteriol. 174, 2281-2287, 1992  
 A>Title: Cloning, characterization, and inactivation of the Bacillus brevis lon gene.  
 A/Reference number: A42375; MUID:9202157; PMID:1551846  
 A/Accession: A42375  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-52 <ITO>  
 A/Cross-references: GB:D00863; NID:g216293; PIDN:BAA00736.1; PID:g303507

Query Match 100.0%; Score 19; DB 2; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 47 VAEF 50

RESULT 5  
 B85745  
 unknown protein encoded within prophage CP-933R [imported] - Escherichia coli (strain O1  
 C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C/Accession: B85745  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A/Reference number: A85480; MUID:21074935; PMID:11206551  
 A/Accession: B85745  
 A/Status: preliminary

```

Molecule type: DNA
;Residues: 1-54 <STO>
;Cross-references: GB:AE005174; NID:gl2515374; PIDN:AAG56425.1; GSPDB:GN00145; UWGP:Z23
;Experimental source: strain O157:H7, substrain EDU933
;Genetics:
;Gene: Z2382

Query Match      100.0%; Score 19; DB 2; Length 54;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/
1 VAEF 4
|||
32 VAEF 35

RESULT 6
;Species: Archaeoglobus fulgidus
;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jan-2000
;Accession: A63386
;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
; ture 390, 364-370, 1997
;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.
; ith, H.O.; Woese, C.R.; Venter, J.C.
;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
;Title reference number: A69250; MUID:98049343; PMID:9389475
;Accession: A63386
;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-59 <KLE>
;Cross-references: GB:AE001028; GB:AE000782; NID:g2689351; PIDN:AAB90157.1; PID:g264950
;Superfamily: Methanococcus jannaschii hypothetical protein MJ0975

Query Match      100.0%; Score 19; DB 2; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/
1 VAEF 4
|||
51 VAEF 54

RESULT 7
;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
;Accession: E71349
;Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
; son, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
; ley, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
; nience 281, 375-388, 1998
;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
;Reference number: A71250; MUID:98332770; PMID:9665876
;Accession: E71349
;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-59 <COL>
;Cross-references: GB:AE001205; GB:AE000520; NID:g3322501; PIDN:AAG65223.1; PID:g332250
;Experimental source: strain Nichols
;Genetics:
;Gene: TP0235

Query Match      100.0%; Score 19; DB 2; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/
1 VAEF 4
|||

Molecule type: DNA
;Residues: 1-63 <ARI>
;Cross-references: DDBJ:D30759; NID:g551628; PIDN:BAA06419.1; PID:d1006989; PID:g551629
;A note: The source was designated as Streptomyces thermotolerans
;C:Genetics:
;A:Gene: carA
;C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
;C:Keywords: ATP
;P:1-55/Domain: ATP-binding cassette homology (fragment) <ABC2>

Query Match      100.0%; Score 19; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 VAEF 4
|||
57 VAEF 60

Db

RESULT 8
;Species: Streptomyces sp.
;C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 20-Sep-1999
;Accession: J04002
;R:Arisawa, A.; Tsunekawa, H.; Okamura, K.; Okamoto, R.
; Biosci. Biotechnol. Biochem. 59, 582-588, 1995
;A:Title: Nucleotide sequence analysis of the carbomycin biosynthetic genes including the
;A:Reference number: J04001; MUID:95290751; PMID:7772821
;A:Accession: J04002
;A:Molecule type: DNA
;A:Residues: 1-63 <ARI>
;A:Cross-references: DDBJ:D30759; NID:g551628; PIDN:BAA06419.1; PID:d1006989; PID:g551629
;A note: The source was designated as Streptomyces thermotolerans
;C:Genetics:
;A:Gene: carA
;C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
;C:Keywords: ATP
;P:1-55/Domain: ATP-binding cassette homology (fragment) <ABC2>

Query Match      100.0%; Score 19; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 VAEF 4
|||
57 VAEF 60

Db

RESULT 9
;Species: Haemophilus influenzae
;C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
;Accession: G64007
;R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
; Science 269, 496-512, 1995
;A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.
; A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
;A:Reference number: A64000; MUID:95350630; PMID:7542800
;A:Accession: G64007
;A:Status: nucleic acid sequence not shown; translation not shown
;A:Molecule type: DNA
;A:Residues: 1-63 <TIGR>
;A:Cross-references: GB:U32728; GB:L42023; NID:gl573425; PIDN:AAC22109.1; PID:gl573437; T]

Query Match      100.0%; Score 19; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 VAEF 4
|||
27 VAEF 30

Db

RESULT 10
;Species: Salmo salar (Atlantic salmon)
;C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
;Accession: I51350
;R:Kvingedal, A.M.
; Gene 150, 335-339, 1994
;A:Title: Characterization of the 5' region of the Atlantic salmon (Salmo salar) transferri
;A:Reference number: I51350; MUID:95121925; PMID:7821802
;A:Accession: I51350
;A:Status: preliminary; translated from GB/ENBL/DBUJ

```

A:Molecule type: DNA  
 A:Residues: 1-64 <KVI>  
 A:Cross-references: GB:L26909; NID:G598395; PIDN:AAC42221.1; PID:G598396  
 C:Genetics:  
 A:Gene: Tf  
 A:Introns: 14/1  
 C:Superfamily: transferrin; transferrin repeat homology

Query Match 100.0%; Score 19; DB 2; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 45 VAEF 48

## RESULT 11

hypothetical protein CAC1179 [imported] - Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C:Accession: D97045  
 R:Noelling, J.; Brston, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
 A:Reference number: A96900; MUID:21359325; PMID:121359325  
 A:Accession: D97045  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-74 <KUR>  
 A:Cross-references: GB:AF001437; PIDN:AAK79151.1; PID:gl5024100; GSPDB:GNO0168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC1179

Query Match 100.0%; Score 19; DB 2; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 5 VAEF 8

## RESULT 12

Gene D protein - phage 186 (fragment)  
 C:Species: phage 186  
 C:Date: 28-May-1993 #sequence\_revision 28-May-1993 #text\_change 17-Mar-1999  
 C:Accession: S07277; S25273  
 R:Kallionis, B.; Dodd, I. B.; Egan, J. B.  
 J. Mol. Biol. 191, 193-209, 1986  
 A:Title: Control of gene expression in the P2-related template coliphages. III. DNA sequence of the P2-related template coliphage 186. X. The cl repressor  
 A:Reference number: S07277; MUID:87112711; PMID:3806670  
 A:Accession: S07277  
 A:Molecule type: DNA  
 A:Residues: 1-75 <KAL>  
 A:Cross-references: ENBL:X04449; NID:gl5549; PID:gl5550  
 R:Dibbens, J.A.; Gregory, S.L.; Egan, J.B.  
 Mol. Microbiol. 6, 2643-2650, 1992  
 A:Title: Control of gene expression in the temperate coliphage 186. X. The cl repressor  
 A:Reference number: S25273; MUID:93078618; PMID:1447973  
 A:Accession: S25273  
 A:Molecule type: DNA  
 A:Residues: 1-75 <DIB>  
 A:Cross-references: ENBL:X04449; NID:gl5549; PID:gl5550

Query Match 100.0%; Score 19; DB 2; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 5 VAEF 8

## RESULT 13

hypothetical protein A532L - Chlorella virus PBCV-1  
 C:Species: Chlorella virus PBCV-1  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T18034  
 R:Graves, M.V.; Van Etten, J.L.  
 submitted to the EMBL Data Library, May 1999  
 A:Reference number: Z18806  
 A:Accession: T18034  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-79 <GRA>  
 A:Cross-references: ENBL:U42580; NID:G4028896; PIDN:AAC96899.1  
 A:Experimental source: specific host Chlorella strain NC64A  
 C:Genetics:  
 A:Note: A532L

Query Match 100.0%; Score 19; DB 2; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 49 VAEF 52

## RESULT 14

hypothetical protein ZC477.6 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T27605  
 R:Du, Z.  
 submitted to the EMBL Data Library, November 1995  
 A:Description: The sequence of C. elegans cosmid ZC477.  
 A:Reference number: Z20392  
 A:Accession: T27605  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-79 <DUZ>  
 A:Cross-references: ENBL:U40802; PIDN:AAA81506.1; CESP:ZC477.6  
 C:Genetics:  
 A:Gene: CESP:ZC477.6  
 A:Introns: 51/2

Query Match 100.0%; Score 19; DB 2; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 74 VAEF 77

## RESULT 15

hypothetical protein BH2192 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: F83922  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: F83922  
 A:Status: preliminary  
 A:Molecule type: DNA

```

;Residues: 1-84 <STO>
;Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05901.1; GSPDB:GN00
;Experimental source: strain C-125
;Genetics:
;Gene: BH2182

Query Match          100.0%; Score 19; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4
   ||||
b 38 VAEF 41

RESULT 16
38725
transferrin - sheep (fragments)
;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
;Date: 20-Sep-1991 #sequence_revision 20-Sep-1991 #text_change 05-Dec-1997
;Accession: A38725
;Tu, G.F.; Achen, M.G.; Aldred, A.R.; Southwell, B.R.; Schreiber, G.
. Biol. Chem. 266, 6201-6208, 1991
;Title: The distribution of cerebral expression of the transferrin gene in species sp
;Reference number: A38725; MUID:91177867; PMID:1848850
;Accession: A38725
;Status: preliminary
;Molecule type: mRNA
;Residues: 1-87 <TUA>
;Cross-references: GB:M64692
;Superfamily: transferrin; transferrin repeat homology
;Keywords: duplication

Query Match          100.0%; Score 19; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4
   ||||
b 13 VAEF 16

RESULT 17
34116
ribosomal protein S15 - Haemophilus influenzae (strain Rd KW20)
;Species: Haemophilus influenzae
;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
;Accession: H64116; E64125
;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.P.; Kerlavage, A.
;Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
;D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
;Science 269, 496-512, 1995
;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
;Reference number: A64000; MUID:93350630; PMID:7542800
;Accession: H64116
;Status: nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-89 <TIG1>
;Cross-references: GB:U32812; GB:L42023; NID:g1574784; PIDN:AAC22973.1; PID:g1574788; T
;Note: named as homolog to a protein from Escherichia coli
;Accession: E64125
;Status: nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-89 <TIG2>
;Cross-references: GB:U32825; GB:L42023; NID:g1574307; PIDN:AAC23117.1; PID:g1574309; T
;Note: named by homology to a protein from Escherichia coli
;Genetics:
;Note: Two copies of this gene are found in the Haemophilus influenzae chromosome
;Superfamily: Escherichia coli ribosomal protein S15; eubacterial ribosomal protein S15
;23-89/Domain: eubacterial ribosomal protein S15 homology <BS15>

Query Match          100.0%; Score 19; DB 1; Length 89;

```

```

Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
   ||||
DB 12 VAEF 15

RESULT 18
F86818
hypothetical protein YpJB [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
;Species: Lactococcus lactis subsp. lactis
;CDate: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
;CAccession: F86818
;R;Bolotin, A.; Wincker, P.; Manger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich
;Genome Res. 11, 731-753, 2001
;A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssi
;Reference number: A86625; MUID:21235186; PMID:11337471
;Accession: F86818
;A;Status: preliminary
;A;Molecule type: DNA
;A;Residues: 1-89 <STO>
;A;Cross-references: GB:AE005176; PID:g12724551; PIDN:AAK05648.1; GSPDB:GN00146
;A;Experimental source: strain IL1403
;C;Genetics:
;A;Gene: YpJB

Query Match          100.0%; Score 19; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
   ||||
DB 68 VAEF 71

RESULT 19
T15018
hypothetical protein Y1103 - Yersinia pestis plasmid pMT1
;Species: Yersinia pestis
;CDate: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
;CAccession: T15018
;Rindler, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.
;Infect. Immun. 66, 5731-5742, 1998
;A;Title: Complete DNA sequence and detailed analysis of the Yersinia pestis KIM5 plasmid
;Reference number: Z18268; MUID:99043898; PMID:9826348
;Accession: T15018
;A;Status: preliminary; translated from GB/EMBL/DDBJ
;A;Molecule type: DNA
;A;Residues: 1-89 <LIN>
;A;Cross-references: EMBL:AF074611; NID:g3883003; PID:g3883101; PIDN:AAC82761.1
;C;Genetics:
;A;Gene: Y1103
;A;Genome: plasmid pMT1

Query Match          100.0%; Score 19; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
   ||||
DB 3 VAEF 6

RESULT 20
AB3053
hypothetical protein Atu4039 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
;Species: Agrobacterium tumefaciens
;CDate: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
;CAccession: AB3053
;R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
;erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClellan
; Karp, P.; Romero, P.; Zhang, S.

```

Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.E.W.  
A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
A:Reference number: AB2577; MUID:21609550; PMID:11743193  
A:Accession: AB3053  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-96 <KUR>  
A:Cross-references: GB:AE008689; PIDN:AA444840.1; PID:gl7742484; GSPDB:GNO0187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu4039  
A:Map position: linear chromosome

Query Match 100.0%; Score 19; DB 2; Length 96;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|  
|  
|  
|  
55 VAEF 58

Db

RESULT 21  
AC1397  
conserved hypothetical protein lmo2579 [imported] - *Listeria monocytogenes* (strain EGD-e)  
C:Species: *Listeria monocytogenes*  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AC1397  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Etian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; McKee, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AC1397  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-97 <GLA>  
A:Cross-references: GB:NC 003210; PIDN:CAD00657.1; PID:gl6412067; GSPDB:GNO0177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo2579

Query Match 100.0%; Score 19; DB 2; Length 97;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|  
|  
|  
|  
15 VAEF 18

Db

RESULT 22  
3B3244  
conserved hypothetical protein PA3202 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
C:Species: *Pseudomonas aeruginosa*  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E83244  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brice, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: E83244  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-99 <STO>  
A:Cross-references: GB:AE004744; GB:AE004091; NID:g9949317; PIDN:AAG06590.1; GSPDB:GNO01

A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA3202

Query Match 100.0%; Score 19; DB 2; Length 99;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|  
|  
|  
|  
62 VAEF 65

Db

RESULT 23  
H90297  
conserved hypothetical protein [imported] - *Sulfolobus solfataricus*  
C:Species: *Sulfolobus solfataricus*  
C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: H90297  
R:She, Q.; Singh, R.K.; Confalonieri, P.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.; arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: *Sulfolobus solfataricus* complete genome.  
A:Reference number: A99139  
A:Accession: H90297  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-101 <KUR>  
A:Cross-references: GB:AE006641; NID:gl3814625; PIDN:AAK41639.1; GSPDB:GNO0155  
C:Genetics:  
A:Gene: SS01404

Query Match 100.0%; Score 19; DB 2; Length 101;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|  
|  
|  
|  
20 VAEF 23

Db

RESULT 24  
TI0097  
nifX protein - *Methanococcus maripaludis*  
C:Species: *Methanococcus maripaludis*  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: TI0097  
R:Kessler, P.S.; McLarnan, J.; Leigh, J.A.  
J. Bacteriol. 179, 541-543, 1997  
A:Title: Nitrogenase phylogeny and the molybdenum dependence of nitrogen fixation in *Met*  
A:Reference number: Z16944; MUID:97144542; PMID:8990309  
A:Accession: TI0097  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-106 <RES>  
A:Cross-references: EMBL:U75887; NID:gl666882; PIDN:AAC45519.1; PID:gl666890  
A:Experimental source: strain JJ  
C:Genetics:  
A:Gene: nifX

Query Match 100.0%; Score 19; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|  
|  
|  
|  
35 VAEF 38

Db

RESULT 25  
C64490  
hypothetical protein MJ1524 - *Methanococcus jannaschii*

```

;Species: Methanococcus jannaschii
;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 12-Jun-2003
;Accession: C64490
;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
;Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
;son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
;Science 273, 1058-1073, 1996
;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
;Reference number: A64300; MUID:96337399; PMID:8688087
;Accession: C64490
;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-108 <BUL>
;Cross-references: GB:U67593; GB:L77117; NID:G2826427; PIDN:AAB99550.1; PID:G1592154; T
;Genetics:
;Map position: FOR1501406-1501732
;Superfamily: uncharacterized conserved protein MJ1524

Query Match      100.0%; Score 19; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      1 VAEF 4
      ||||
b     52 VAEF 55

RESULT 26
70949
robable regulator - Mycobacterium tuberculosis (strain H37RV)
;Species: Mycobacterium tuberculosis
;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
;Accession: E70949
;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
;Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.
;Andream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.
;Science 283, 537-544, 1998
;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
;Reference number: A70500; MUID:98295987; PMID:9634230
;Accession: E70949
;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-109 <COL>
;Cross-references: GB:AL021646; GB:AL123456; NID:G3242278; PIDN:CAAL16648.1; PID:G282759
;Experimental source: strain H37RV
;Genetics:
;Gene: RV3183

Query Match      100.0%; Score 19; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      1 VAEF 4
      ||||
b     97 VAEF 100

RESULT 27
17582
vpothetical protein A92L - Chlorella virus PBCV-1
;Species: Chlorella virus PBCV-1
;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
;Accession: T17582
;Graves, M.V.; Van Etten, J.L.
;submitted to the EMBL Data Library, May 1999
;Reference number: Z18806
;Accession: T17582
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-111 <GRA>
;Cross-references: EMBL:U42580; NID:G4028896; PIDN:AAC96460.1

;Species: Methanococcus jannaschii
;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 12-Jun-2003
;Accession: C64490
;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
;Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
;son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
;Science 273, 1058-1073, 1996
;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
;Reference number: A64300; MUID:96337399; PMID:8688087
;Accession: C64490
;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-108 <BUL>
;Cross-references: GB:U67593; GB:L77117; NID:G2826427; PIDN:AAB99550.1; PID:G1592154; T
;Genetics:
;Map position: FOR1501406-1501732
;Superfamily: uncharacterized conserved protein MJ1524

Query Match      100.0%; Score 19; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      1 VAEF 4
      ||||
b     52 VAEF 55

RESULT 26
70949
robable regulator - Mycobacterium tuberculosis (strain H37RV)
;Species: Mycobacterium tuberculosis
;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
;Accession: E70949
;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
;Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.
;Andream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.
;Science 283, 537-544, 1998
;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
;Reference number: A70500; MUID:98295987; PMID:9634230
;Accession: E70949
;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-109 <COL>
;Cross-references: GB:AL021646; GB:AL123456; NID:G3242278; PIDN:CAAL16648.1; PID:G282759
;Experimental source: strain H37RV
;Genetics:
;Gene: RV3183

Query Match      100.0%; Score 19; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      1 VAEF 4
      ||||
b     97 VAEF 100

RESULT 27
17582
vpothetical protein A92L - Chlorella virus PBCV-1
;Species: Chlorella virus PBCV-1
;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
;Accession: T17582
;Graves, M.V.; Van Etten, J.L.
;submitted to the EMBL Data Library, May 1999
;Reference number: Z18806
;Accession: T17582
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-111 <GRA>
;Cross-references: EMBL:U42580; NID:G4028896; PIDN:AAC96460.1

;Species: Pseudomonas aeruginosa
;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
;Accession: B83431
;R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
;adhan, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
;Lory, S.; Olson, M.V.
;Nature 406, 959-964, 2000
;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
;Reference number: A82950; MUID:20437337; PMID:10984043
;Accession: B83431
;Status: preliminary
;Molecule type: DNA
;Residues: 1-112 <STO>
;Cross-references: GB:AE004598; GB:AE004091; NID:G9947687; PIDN:AAG05111.1; GSPDB:GN001
;Experimental source: strain PA01
;Genetics:
;Gene: psci; PA1722
;Superfamily: Versinia enterocolitica plasmid pYV virC-region hypothetical protein yscI

Query Match      100.0%; Score 19; DB 2; Length 112;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      1 VAEF 4
      ||||
b     28 VAEF 31

RESULT 29
A55872
kedarcidin [validated] - Streptoalloteichus sp. (ATCC 53650)
;Alternate names: kedarcidin apoprotein
;Species: Streptoalloteichus sp.
;Variety: strain L585-6 (ATCC 53650)
;Date: 21-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 15-Sep-2000
;Accession: A55872
;R;Hofstead, S.J.; Watson, J.A.; Malacko, A.R.; Marquardt, H.
;J. Antibiot. 45, 1250-1254, 1992
;Title: Kedarcidin, a new chromoprotein antitumor antibiotic. II. Isolation, purificati
;Reference number: A55872; MUID:93015257; PMID:1399845
;Accession: A55872
;Molecule type: protein
;Residues: 1-114 <HOF>
;Note: sequence extracted from NCBI backbone (NCBIP:118354)
;R;Constantine, K.L.; Colson, K.L.; Wittekind, M.; Friedrichs, M.S.; Zein, N.; Tuttle, J.,
;et al.
;submitted to the Brookhaven Protein Data Bank, June 1994
;Reference number: A52519; PDB:1AKP
;Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR, residues
;R;Zein, N.; Casazza, A.M.; Doyle, I.W.; Leet, J.E.; Schroeder, D.R.; Solomon, W.; Nadler,
;Proc. Natl. Acad. Sci. U.S.A. 90, 8009-8012, 1993
;Title: Selective proteolytic activity of the antitumor agent kedarcidin.
;Reference number: A58601; MUID:93376732; PMID:8367457
;Contents: annotation; protein activity
;Note: the apoprotein may have proteolytic activity
;Superfamily: macromycinin

```

C;Keywords: antibiotic; pigment binding  
F:37-47,98-95/disulfide bonds: #status experimental

Query Match 100.0%; Score 19; DB 2; Length 114;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1 VAEF 4  
|||  
49 VAEF 52

2B

RESULT 30  
D97414  
hypothetical protein AGR\_C 804 [imported] - Agrobacterium tumefaciens (strain C58, Cerec  
C;Species: Agrobacterium tumefaciens  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C;Accession: D97414  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: D97414  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-114 <KUR>  
A;Cross-references: GB:AE007869; PIDN:AAK86269.1; PID:g15155379; GSPDB:GN00169  
C;Genetics:  
A;Gene: AGR\_C 804  
A;Map position: circular chromosome

Query Match 100.0%; Score 19; DB 2; Length 114;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1 VAEF 4  
|||  
40 VAEF 43

2B

RESULT 31  
D75218  
hypothetical protein PAB0185 - Pyrococcus abyssi (strain Orsay)  
C;Species: Pyrococcus abyssi  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C;Accession: D75218  
R;Anonymous, Genoscope  
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru  
submitted to the EMBL Data Library, July 1999  
A;Reference number: A75001  
A;Accession: D75218  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-115 <KAW>  
A;Cross-references: GB:AJ248283; GB:AL096836; NID:G5457433; PIDN:CAB49195.1; PID:ei51508  
A;Experimental source: strain Orsay  
C;Genetics:  
A;Gene: PAB0185

Query Match 100.0%; Score 19; DB 2; Length 115;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||  
101 VAEF 104

2B

RESULT 32  
IPAF  
insulin precursor - American goosfish  
C;Species: Lophius americanus (American goosfish)

C;Date: 31-Mar-1981 #sequence\_revision 31-Mar-1981 #text\_change 16-Jul-1999  
C;Accession: A01608  
R;Hobart, P.M.; Shen, L.P.; Crawford, R.; Pictet, R.L.; Rutter, W.J.  
Science 210, 1360-1363, 1980  
A;Title: Comparison of the nucleic acid sequence of anglerfish and mammalian insulin mRN  
A;Reference number: A94254; MUID:81056434; PMID:7001633  
A;Accession: A01608  
A;Molecule type: mRNA  
A;Residues: 1-116 <HOB>  
A;Cross-references: GB:V00634; GB:J00934; NID:g64025; PIDN:CAA23907.1; PID:g64026  
C;Superfamily: insulin  
C;Keywords: hormone; pancreas  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-54/Domain: insulin chain B #status predicted <BC>  
F:25-54/Domain: insulin #status predicted <MAT>  
F:56-93/Domain: connecting peptide #status predicted <CPEP>  
F:96-116/Domain: insulin chain A #status predicted <ACH>  
F:132-102,44-115,101-106/Disulfide bonds: #status predicted

Query Match 100.0%; Score 19; DB 1; Length 116;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||  
80 VAEF 83

2B

RESULT 33  
T48682  
hypothetical protein DKFZp761N05121.1 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 05-May-2000  
C;Accession: T48682  
R;Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z24535  
A;Accession: T48682  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-118 <AAA>  
A;Cross-references: EMBL:AL353932  
A;Experimental source: adult amygdala; clone DKFZp761N05121  
C;Genetics:  
A;Note: DKFZp761N05121.1

Query Match 100.0%; Score 19; DB 2; Length 118;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||  
15 VAEF 18

2B

RESULT 34  
F95327  
hypothetical protein SMA0974 [imported] - Sinorhizobium meliloti (strain 1021) magaplaem  
C;Species: Sinorhizobium meliloti  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C;Accession: F95327  
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows  
; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C  
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilot  
A;Reference number: A95262; MUID:31396509; PMID:11481432  
A;Accession: F95327  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-119 <KUR>  
A;Cross-references: GB:AE006469; PIDN:AAK65184.1; PID:g14523629; GSPDB:GN00165  
A;Experimental source: strain 1021, megaplaemid pSymA  
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

```

ela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Pederspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, I.
Science 293, 668-672, 2001
;Authors: Kahn, D.; Kahn, M.L.; Kaiman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
;ebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
;Reference number: A96039; MUID:21368234; PMID:11474104
;Contents: annotation
;Genetics:
;Gene: SWA0974
;Genome: plasmid

Query Match 100.0%; Score 19; DB 2; Length 119;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4
|||
26 VAEF 29

RESULT 35
32861
brnugal transfer protein XFA0006 [imported] - Xylella fastidiosa (strain 9a5c)
;Species: Xylella fastidiosa
;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
;Accession: F82861
;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
;ature 406, 151-157, 2000
;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
;Reference number: A82515; MUID:20365717; PMID:10910347
;Note: for a complete list of authors see reference number A59328 below
;Accession: F82861
;Status: preliminary
;Molecule type: DNA
;Residues: 1-122 <SIM>
;Cross-references: GB:AE003851; NID:G9112238; PIDN:AAE85575.1; GSPDB:GN00130; XFSC:XFAC
;Experimental source: strain 9a5c
;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
;S-neto, B.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
;mitted to GenBank, June 2000
;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
;D.; Junqueira, M.B.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh
;hado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
;F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
;drigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak
;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
;Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
;Reference number: A59328
;Contents: annotation
;Genetics:
;Gene: XFA0006
;Genome: plasmid
;Note: plasmid pXF5.1

Query Match 100.0%; Score 19; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4
|||
32 VAEF 35

RESULT 36
69129
ama-carboxymuconolactone decarboxylase - Methanobacterium thermoautotrophicum (strain
;Species: Methanobacterium thermoautotrophicum
;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Jun-1999
;Accession: B69129
;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;

```

```

;Oiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
;ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct.
;A:Reference number: A69000; MUID:98037514; PMID:9371463
;Accession: B69129
;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-125 <MTH>
;Cross-references: GB:AE000810; GB:AE000666; NID:G2621277; PIDN:AMB84740.1; PID:G262128.
;Experimental source: strain Delta H
;Genetics:
;Gene: MTH234
;Superfamily: 4-carboxymuconolactone decarboxylase

Query Match 100.0%; Score 19; DB 2; Length 125;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4
|||
37 VAEF 40

RESULT 37
D70790
hypothetical protein RV3675 - Mycobacterium tuberculosis (strain H37RV)
;Species: Mycobacterium tuberculosis
;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
;Accession: D70790
;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
;Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
;Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
;Reference number: A70500; MUID:98295987; PMID:9634230
;Accession: D70790
;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-125 <COL>
;Cross-references: GB:AL022121; GB:AL123456; NID:G3261559; PIDN:CAA17997.1; PID:e126453;
;Experimental source: strain H37RV
;Genetics:
;Gene: RV3675

Query Match 100.0%; Score 19; DB 2; Length 125;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4
|||
42 VAEF 45

RESULT 38
B81368
hypothetical protein Cj0939c [imported] - Campylobacter jejuni (strain NCTC 11168)
;Species: Campylobacter jejuni
;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
;Accession: B81368
;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
;C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
;Reference number: A81250; MUID:20150912; PMID:10688204
;Accession: B81368
;Status: preliminary
;Molecule type: DNA
;Residues: 1-126 <PAR>
;Cross-references: GB:AL139076; GB:AL111168; NID:G6968128; PIDN:CAE73196.1; PID:G696837;
;Experimental source: serotype O2, strain NCTC 11168
;Genetics:

```

A.Gene: Cj0399c

Query Match  
Best Local Similarity 100.0%; Score 19; DB 2; Length 126;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 83 VAEF 86

RESULT 39

D70799  
hypothetical protein RV3747 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: D70799  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: D70799  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-127 <COL>  
A:Cross-references: GB:AL022121; GB:AL123456; NID:G3261559; PIDN:CAA18069.1; PID:G296017  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV3747

Query Match  
Best Local Similarity 100.0%; Score 19; DB 2; Length 127;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 80 VAEF 83

RESULT 40

AB1732  
hypothetical protein lin2399 [imported] - Listeria innocua (strain Clip11262)  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AB1732  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma, J.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, J.; Jones, L.M.; Karst, U.  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AB1732  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-127 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CA097626.1; PID:G16414922; GSPDB:GN00178  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: lin2399

Query Match  
Best Local Similarity 100.0%; Score 19; DB 2; Length 127;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 71 VAEF 74

RESULT 41

A54797  
ileal lipid-binding protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 05-Nov-1999  
C:Accession: A54797  
R:Crossman, M.W.; Hautt, S.M.; Gordon, J.I.  
J. Cell Biol. 126, 1547-1564, 1994  
A:Title: The mouse ileal lipid-binding protein gene: a model for studying axial pattern  
A:Reference number: A54797; MUID:94375529; PMID:8089185  
A:Accession: A54797  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-128 <CKO>  
A:Cross-references: GB:U00938; NID:G507143; PIDN:AAC27352.1; PID:G507144  
C:Genetics:  
A:Introns: 23/1; 81/3; 111/3  
C:Superfamily: myelin P2 protein

Query Match  
Best Local Similarity 100.0%; Score 19; DB 2; Length 128;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 92 VAEF 95

RESULT 42

T02012  
probable cathepsin B-like cysteine proteinase (EC 3.4.22.-) T15B16.17b - Arabidopsis th  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 22-Jun-1999  
C:Accession: T02012  
R:Stoneking, T.; Smith, R.  
submitted to the EMBL Data Library, November 1998  
A:Description: The sequence of A. thaliana T15B16.  
A:Reference number: Z14488  
A:Accession: T02012  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-129 <STO>  
A:Cross-references: EMBL:AF104919; NID:G3859590; PIDN:AA072873.1; PID:G38595607  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 26/3; 44/3; 71/3  
A:Note: T15B16.17b  
C:Superfamily: papain  
C:Keywords: cysteine proteinase; hydrolase

Query Match  
Best Local Similarity 100.0%; Score 19; DB 2; Length 129;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 72 VAEF 75

RESULT 43

D72348  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: D72348  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571

```

Accession: D72348
Status: preliminary
Molecule type: DNA
Residues: 1-130 <ARN>
Cross-references: GB:AB001739; GB:AB000512; NID:G4981176; PIDN:RAD35739.1; PID:G498117
Experimental source: strain MSB8
Gene: TM0655
Superfamily: conserved hypothetical protein M00315

Query Match      100.0%; Score 19; DB 2; Length 130;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/      1 VAEF 4
      ||||
      9 VAEF 12

RESULT 44
75208
Hypothetical protein PAB2224 - Pyrococcus abyssi (strain Orsay)
Species: Pyrococcus abyssi
Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 06-Jan-2003
Accession: G75208
anonymous, Genoscope
Submitted to the EMBL Data Library, July 1999
Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
Reference number: A75001
Accession: G75208
Status: preliminary
Molecule type: DNA
Residues: 1-131 <XAW>
Cross-references: GB:AJ248283; GB:AL096836; NID:G5457433; PIDN:CAB49118.1; PID:G545762
Experimental source: strain Orsay
Genetics:
Gene: PAB2224
Superfamily: uncharacterized conserved protein

Query Match      100.0%; Score 19; DB 2; Length 131;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y      1 VAEF 4
      ||||
      4 VAEF 7

RESULT 45
Q2270
Hypothetical 15.2K protein - Anabaena sp.
Species: Anabaena sp.
Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 13-Sep-1998
Accession: JQ2270
Larimer, F.W.; Soper, T.S.
ene 126, 85-92, 1993
Title: Overproduction of Anabaena 7120 ribulose-bisphosphate carboxylase/oxygenase in
Reference number: JQ2270; MUID:93231541; PMID:8472962
Accession: JQ2270
Molecule type: DNA
Residues: 1-132 <LAR>
Cross-references: DDBJ:J01540
Experimental source: strain 7120
Genetics:
Gene: rbcX
Superfamily: rbcX protein

Query Match      100.0%; Score 19; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y      1 VAEF 4
      ||||

Accession: D72348
Status: preliminary
Molecule type: DNA
Residues: 1-130 <ARN>
Cross-references: GB:AB001739; GB:AB000512; NID:G4981176; PIDN:RAD35739.1; PID:G498117
Experimental source: strain MSB8
Gene: TM0655
Superfamily: conserved hypothetical protein M00315

Query Match      100.0%; Score 19; DB 2; Length 130;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/      1 VAEF 4
      ||||
      9 VAEF 12

RESULT 46
AG1996
Hypothetical protein alr1525 [imported] - Nostoc sp. (strain PCC 7120)
Species: Nostoc sp. PCC 7120
Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
Accession: AG1996
R.Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shampo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anat
AReference number: AB1807; MUID:21595285; PMID:11759840
AAccession: AG1996
AStatus: preliminary
A Molecule type: DNA
AResidues: 1-132 <KUR>
ACross-references: GB:BA000019; PIDN:BA877891.1; PID:G17135345; GSPDB:GN00179
AExperimental source: strain PCC 7120
Genetics:
AGene: alr1525
Superfamily: rbcX protein

Query Match      100.0%; Score 19; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VAEF 4
      ||||
      81 VAEF 84

Db

RESULT 47
AH0137
conserved hypothetical protein YPO1120 [imported] - Yersinia pestis (strain CO92)
Species: Yersinia pestis
Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
Accession: AH0137
R.Parthill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; I
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, E
Nature 413, 523-527, 2001
A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.
AReference number: AB0001; MUID:21470413; PMID:11586360
Accession: AH0137
Status: preliminary
A Molecule type: DNA
Residues: 1-133 <KUR>
Cross-references: GB:AL590842; PIDN:CAC89963.1; PID:G15979187; GSPDB:GN00175
Genetics:
AGene: YPO1120
Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match      100.0%; Score 19; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VAEF 4
      ||||
      129 VAEF 132

Db

RESULT 48
H75623
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
Species: Deinococcus radiodurans
Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
Accession: H75623
R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Sher, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

```

Science 286, 1571-1577, 1999  
 A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896; PMID:10567266  
 A:Accession: H75623  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-133 <WHI>  
 A:Cross-references: GB:AE001826; NID:g6460827; PIDN:AAF12590.1; PID:g6460886; TIGR:DRB00  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DRB0052  
 A:Map position: megaplasmid  
 A:Genome: plasmid  
 A>Note: plasmid MPI

Query Match 100.0%; Score 19; DB 2; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 18 VAEF 21

RESULT 49  
 A4221  
 A>Title: riboflavin synthase beta subunit [imported] - Halobacterium sp. NRC-1  
 A:Species: Halobacterium sp. NRC-1  
 A:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 A:Accession: A84221  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-133 <STO>  
 A:Cross-references: GB:A84160; MUID:20504483; PMID:11016950  
 A:Reference number: A84160; MUID:20504483; PMID:11016950  
 A:Accession: A84221  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-133 <STO>  
 A:Cross-references: GB:AB004437; NID:g10580221; PIDN:AAG19133.1; GSPDB:GN00138  
 A:Experimental source: strain A84160; MUID:20504483; PMID:11016950  
 C:Genetics:  
 A:Gene: ribE

Query Match 100.0%; Score 19; DB 2; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 8 VAEF 11

RESULT 50  
 WMEC15  
 A>Title: ybgC protein - Escherichia coli (strain K-12)  
 A:Species: Escherichia coli  
 A:Date: 15-Oct-1982 #sequence\_revision 30-Jun-1988 #text\_change 01-Mar-2002  
 A:Accession: A25980; G64809  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-134 <SUN>  
 A:Cross-references: GB:M16499; NID:g148021; PIDN:AAA83918.1; PID:g1128977  
 A:Experimental source: strain K-12  
 C:Genetics:  
 A:Gene: ybgC

Query Match 100.0%; Score 19; DB 2; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 8 VAEF 11

RESULT 51  
 C90725  
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: C90725  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-134 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA834194.1; PID:g13360230; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: ECs0771  
 A:Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match 100.0%; Score 19; DB 2; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 129 VAEF 132

RESULT 52  
 D85576  
 A>Title: ybgC protein - Escherichia coli (strain O157:H7, substrain EDL93  
 A:Species: Escherichia coli  
 A:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 A:Accession: D85576  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-134 <STO>  
 A:Cross-references: GB:AE005174; NID:g12513669; PIDN:AAG55072.1; GSPDB:GN00145; UWGP:Z09  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: ybgC

Query Match 100.0%; Score 19; DB 2; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 129 VAEF 132

A:Accession: G64809  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-134 <BLAT>  
 A:Cross-references: GB:AE000177; GB:U00096; NID:g1786955; PIDN:AAC73830.1; PID:g1786957;  
 A:Experimental source: strain K-12, substrain MG1655  
 A:Comment: This is one of the proteins, encoded by the fli-tolAB gene cluster, that is i  
 C:Genetics:  
 A:Gene: ybgC  
 A:Map position: 17 min  
 A:Start codon: GTG  
 A:Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match 100.0%; Score 19; DB 1; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 129 VAEF 132

Query Match 100.0%; Score 19; DB 1; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 129 VAEF 132

RESULT 51  
 C90725  
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: C90725  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-134 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA834194.1; PID:g13360230; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: ECs0771  
 A:Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match 100.0%; Score 19; DB 2; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 129 VAEF 132

hypothetical protein ECs0771 [imported] - Escherichia coli (strain O157:H7, substrain RI  
 A:Species: Escherichia coli  
 A:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 A:Accession: C90725  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-134 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA834194.1; PID:g13360230; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: ECs0771  
 A:Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match 100.0%; Score 19; DB 2; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 129 VAEF 132

RESULT 52  
 D85576  
 A>Title: ybgC protein - Escherichia coli (strain O157:H7, substrain EDL93  
 A:Species: Escherichia coli  
 A:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 A:Accession: D85576  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-134 <STO>  
 A:Cross-references: GB:AE005174; NID:g12513669; PIDN:AAG55072.1; GSPDB:GN00145; UWGP:Z09  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: ybgC

Query Match 100.0%; Score 19; DB 2; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 DB 129 VAEF 132

```

Query Match      100.0%; Score 19; DB 2; Length 134;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4
  ||||
  129 VAEF 132

RESULT 53
D0592
Conserved hypothetical protein STY0790 [imported] - Salmonella enterica subsp. enterica
Species: Salmonella enterica subsp. enterica serovar Typhi
Note: this species has also been called Salmonella typhi
Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
Accession: AD0592
Parkhill, J.; Dougan, G.; James, K.D.; Thomeon, N.R.; Pickard, D.; Wain, J.; Churcher,
S.; Moulie, S.; O'Gaora, P.
ature 413, 848-852, 2001
Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
Reference number: AB0502; MUID:21534947; PMID:11677608
Accession: AD0592
Status: preliminary
Molecule type: DNA
Residues: 1-134 <PAR>
Cross-references: GB:AL513382; PIDN:CAD05206.1; PID:g16501976; GSPDB:GN00176
Gene: STY0790
Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match      100.0%; Score 19; DB 2; Length 134;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4
  ||||
  129 VAEF 132

RESULT 54
76416
Hypothetical protein - Synecocystis sp. (strain PCC 6803)
Species: Synecocystis sp.
Variety: PCC 6803
Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
Accession: S76416
Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
NA Res. 3, 109-136, 1996
Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
Reference number: S74322; MUID:97061201; PMID:8905231
Accession: S76416
Status: preliminary
Molecule type: DNA
Residues: 1-136 <KAN>
Cross-references: EMBL:D90915; GB:AB001339; NID:g1653504; PIDN:BAAL8545.1; PID:g165363
Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match      100.0%; Score 19; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4
  ||||
  88 VAEF 91

RESULT 55
81355
Hypothetical protein Cj0829c [imported] - Campylobacter jejuni (strain NCTC 11168)
Species: Campylobacter jejuni
CDate: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
Accession: E81355
Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
AReference number: AB1250; MUID:20150912; PMID:10688204
Accession: E81355
Status: preliminary
Molecule type: DNA
Residues: 1-137 <PAR>
Cross-references: GB:AL139076; GB:AL111168; NID:g96968128; PIDN:CAB73094.1; PID:g696827;
AExperimental source: serotype O2, strain NCTC 11168
Genetics:
Gene: Cj0829c
Superfamily: hypothetical protein yneT

Query Match      100.0%; Score 19; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
  ||||
  33 VAEF 36

RESULT 56
F82106
Flagellar basal-body rod protein FlgC VC2199 [imported] - Vibrio cholerae (strain N16961)
Species: Vibrio cholerae
CDate: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
Accession: F82106
Reidberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
AReference number: AB2035; MUID:20406833; PMID:10952301
Accession: F82106
Status: preliminary
Molecule type: DNA
Residues: 1-138 <HEI>
Cross-references: GB:AE003852; GB:AE004292; NID:g9656760; PIDN:AAF95344.1; GSPDB:GN001;
AExperimental source: serogroup O1, strain N16961; biotype E1 Tor
Genetics:
Gene: VC2199
Map position: 1

Query Match      100.0%; Score 19; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
  ||||
  79 VAEF 82

RESULT 57
A69052
Riboflavin synthase beta subunit - Methanobacterium thermoautotrophicum (strain Delta H)
Species: Methanobacterium thermoautotrophicum
CDate: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-May-2000
Accession: A69052
Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; F
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwni, N.,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functi
AReference number: A69000; MUID:98037514; PMID:9371463
Accession: A69052
Status: preliminary; nucleic acid sequence not shown; translation not shown
Molecule type: DNA

```

```

>Residues: 1-139 <MTH>
>Cross-references: GB:AE000902; GB:AE000666; NID:g2622500; PIDN:AAB85867.1; PID:g262250
>Experimental source: strain Delta H
>Genetics:
>Gene: MTH1390
>Superfamily: riboflavin synthase beta chain

Query Match      100.0%; Score 19; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 VAEF 4
   ||||
b 10 VAEF 13

RESULT 58
83975
>Hypothetical protein BH2601 [imported] - Bacillus halodurans (strain C-125)
>Species: Bacillus halodurans
>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
>Accession: A83975
>Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
>Nucleic Acids Res. 28, 4317-4331, 2000
>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
>Reference number: A83650; MUID:20512582; PMID:11058132
>Accession: A83975
>Status: Preliminary
>Molecule type: DNA
>Residues: 1-139 <STO>
>Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA806320.1; GSPDB:GN00
>Experimental source: strain C-125
>Genetics:
>Gene: BH2601
>Superfamily: Bacillus subtilis hypothetical protein ylbD

Query Match      100.0%; Score 19; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 VAEF 4
   ||||
b 11 VAEF 14

RESULT 59
869323
>Hypothetical protein AF0586 - Archaeoglobus fulgidus
>Species: Archaeoglobus fulgidus
>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Nov-1999
>Accession: B69323
>Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
>Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
>Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
>ature 390, 364-370, 1997
>Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
>Smith, H.O.; Woese, C.R.; Venter, J.C.
>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
>Reference number: A69250; MUID:98049343; PMID:9389475
>Accession: B69323
>Status: preliminary; nucleic acid sequence not shown; translation not shown
>Molecule type: DNA
>Residues: 1-140 <KLE>
>Cross-references: GB:AE001064; GB:AE000782; NID:g2689387; PIDN:AAB90658.1; PID:g265004
>Superfamily: Archaeoglobus fulgidus hypothetical protein AF0586

Query Match      100.0%; Score 19; DB 2; Length 140;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 VAEF 4
   ||||
b 65 VAEF 68

```

## RESULT 60

```

JQ2279
>Low-temperature regulated protein BN115 - rape
>Species: Brassica napus (rape)
>Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 28-May-1999
>Accession: JQ2279
>Weretilnyk, E.; Orr, W.; White, T.C.; Iu, B.; Singh, J.
>Plant Physiol. 101, 171-177, 1993
>Title: Characterization of three related low-temperature-regulated cDNAs from winter E
>Reference number: JQ2279; MUID:94105287; PMID:7904076
>Accession: JQ2279
>Molecule type: mRNA
>Residues: 1-142 <WER>
>Cross-references: GB:S68726; NID:g544693; PIDN:AAB29482.1; PID:g544694
>Comment: This protein is a low-temperature-related protein.

Query Match      100.0%; Score 19; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 VAEF 4
   ||||
b 125 VAEF 128

```

## RESULT 61

```

JQ2280
>Low-temperature regulated protein BN19 - rape
>Species: Brassica napus (rape)
>Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 28-May-1999
>Accession: JQ2280
>Weretilnyk, E.; Orr, W.; White, T.C.; Iu, B.; Singh, J.
>Plant Physiol. 101, 171-177, 1993
>Title: Characterization of three related low-temperature-regulated cDNAs from winter B
>Reference number: JQ2279; MUID:94105287; PMID:7904076
>Accession: JQ2280
>Molecule type: mRNA
>Residues: 1-142 <WER>
>Cross-references: GB:S68879; NID:g544695; PIDN:AAB29483.1; PID:g544696
>Comment: This protein is a low-temperature-related protein.

Query Match      100.0%; Score 19; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 VAEF 4
   ||||
b 125 VAEF 128

```

## RESULT 62

```

H69515
>riboflavin synthase, subunit beta (ribE) homolog - Archaeoglobus fulgidus
>Species: Archaeoglobus fulgidus
>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
>Accession: H69515
>Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
>Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
>Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
>ature 390, 364-370, 1997
>Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
>Smith, H.O.; Woese, C.R.; Venter, J.C.
>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
>Reference number: A69250; MUID:98049343; PMID:9389475
>Accession: H69515
>Status: preliminary; nucleic acid sequence not shown; translation not shown
>Molecule type: DNA
>Residues: 1-143 <KLE>
>Cross-references: GB:AE000957; GB:AE000782; NID:g2689280; PIDN:AAB89124.1; PID:g264840
>Superfamily: riboflavin synthase beta chain

```

```

Query Match      100.0%; Score 19; DB 2; Length 143;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/ 1 VAEF 4
| 1111
\ 10 VAEF 13

RESULT 63
16978
;Title: transposase, truncated section 1 [imported] - Streptomyces coelicolor (fragment)
;Species: Streptomyces coelicolor
;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
;Accession: T36978
;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
;Submitted to the EMBL Data Library, August 1999
;Reference number: Z21618
;Accession: T36978
;Status: preliminary; translated from GB/EMBL/DDBJ
;Molecule type: DNA
;Residues: 1-143 <OLI>
;Cross-references: EMBL:AL109949; PIDN:CAE52892.1; GSPDB:GN00070; SCOEDB:SCU11.07c
;Experimental source: strain A3(2)
;Comment: A complete transposase sequence can be assembled from three adjacent orfs (PI
;Genetics:
;Gene: SCOEDB:SCU11.07c

Query Match      100.0%; Score 19; DB 2; Length 143;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/ 1 VAEF 4
| 1111
\ 113 VAEF 116

RESULT 64
43715
;Title: ribosomal protein L13 [similarity] - Haloarcula marismortui
;Species: Haloarcula marismortui
;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 14-Apr-2003
;Accession: B41715
;Kroemer, W.J.; Arndt, E.
;Biol. Chem. 266, 24573-24579, 1991
;Title: Halobacterial S9 operon. Three ribosomal protein genes are cotranscribed with g
acterium) marismortui.
;Reference number: A41715; MUID:92105119; PMID:1840597
;Accession: B41715
;Status: preliminary
;Molecule type: DNA
;Residues: 1-145 <KRO>
;Cross-references: GB:M76567; NID:g148775; PIDN:AAA73097.1; PID:g148777
;Superfamily: ribosomal protein L13

Query Match      100.0%; Score 19; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/ 1 VAEF 4
| 1111
\ 3 VAEF 6

RESULT 65
77043
;Title: hypothetical protein s110678 - Synechocystis sp. (strain PCC 6803)
;Species: Synechocystis sp.
;Variety: PCC 6803
;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
;Accession: S77043
;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
;K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S77043
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-145 <KAN>
A;Cross-references: EMBL:D64005; GB:AB001339; NID:g1001779; PIDN:BAA10735.1; PID:g100658
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: Synechocystis hypothetical protein s110678

Query Match      100.0%; Score 19; DB 2; Length 145;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
| 1111
Db 35 VAEF 38

RESULT 66
F89837
;Title: conserved hypothetical protein SA0624 [imported] - Staphylococcus aureus (strain N315)
;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: F89837
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; I
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: F89837
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-146 <KUR>
A;Cross-references: GB:BA000018; PID:g13700560; PIDN:BA041857.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA0624

Query Match      100.0%; Score 19; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
| 1111
Db 102 VAEF 105

RESULT 67
C84294
;Title: hypothetical protein Vng1398c [imported] - Halobacterium sp. NRC-1
;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: C84294
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
;Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: C84294
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-147 <STO>
A;Cross-references: GB:AE004437; NID:g10580903; PIDN:AAG19719.1; GSPDB:GN00138
C;Genetics:
A;Gene: VNG1398C
C;Superfamily: hypothetical protein YJCF

```

Query Match 100.0%; Score 19; DB 2; Length 147;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 ||||  
 b 117 VAEF 120

## RESULT 68

34803  
 Hypothetical protein SC2E1.35 - Streptomyces coelicolor  
 Species: Streptomyces coelicolor  
 Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 04-Mar-2000  
 Accession: T34803  
 Murphy, L.; Harris, D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, June 1998  
 Reference number: Z21557  
 Accession: T34803  
 Status: preliminary; translated from GB/EMBL/DDBJ  
 Molecule type: DNA  
 Residues: 1-147 <MUR>  
 Cross-references: EMBL:AL023797; PIDN:CAA19410.1; GSPDB:GN00070; SC0E1.35  
 Experimental source: strain A3(2)  
 Genetics:  
 Gene: SC0E1.35  
 Superfamily: Streptomyces coelicolor hypothetical protein SC2E1.35

Query Match 100.0%; Score 19; DB 2; Length 147;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 ||||  
 b 125 VAEF 123

## RESULT 69

381788  
 conserved hypothetical protein NMA2160 [imported] - Neisseria meningitidis (strain Z2491)  
 Species: Neisseria meningitidis  
 Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
 Accession: E81788  
 Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
 Holroyd, S.; Jajels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
 Nature 404, 502-506, 2000  
 Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
 Reference number: A81775; MUID:20222556; PMID:10761919  
 Accession: E81788  
 Status: preliminary  
 Molecule type: DNA  
 Residues: 1-148 <PAR>  
 Cross-references: GB:AL162758; GB:AL157959; NID:G7380672; PIDN:CAB85372.1; PID:G738078  
 Experimental source: serogroup A, strain Z2491  
 Genetics:  
 Gene: NMA2160  
 Superfamily: hypothetical protein ytw1

Query Match 100.0%; Score 19; DB 2; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 ||||  
 b 77 VAEF 80

## RESULT 70

310655  
 Hypothetical protein X - Pyrococcus woessii (fragment)  
 Species: Pyrococcus woessii  
 Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 18-Jun-1993  
 Accession: S10655

R;Zwickl, P.; Fabry, S.; Bogedain, C.; Haas, A.; Hensel, R.  
 J. Bacteriol. 172, 4329-4338, 1990  
 Title: Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeobacter  
 Scherichia coli.  
 A;Reference number: S10650; MUID:90330536; PMID:2165475  
 Accession: S10655  
 Molecule type: DNA  
 Residues: 1-148 <ZWI>  
 Note: the authors translated the codon GGT for residue 54 as Glu

Query Match 100.0%; Score 19; DB 2; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 ||||  
 b 115 VAEF 118

## RESULT 71

AC1137  
 ribose 5-phosphate isomerase homolog lmo0498 [imported] - Listeria monocytogenes (strain  
 C;Species: Listeria monocytogenes  
 Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
 Accession: AC1137  
 Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Enrian, K.D.; Fsihi, H.  
 ; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madieno, E.; Maitournam, A.; Ma  
 ck, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
 Title: Comparative genomics of Listeria species.  
 Reference number: AB1077; MUID:21537279; PMID:11679669

Y 1 VAEF 4  
 ||||  
 b 124 VAEF 127

AC1137  
 Status: preliminary  
 Molecule type: DNA  
 Residues: 1-149 <GLA>  
 Cross-references: GB:NC\_003210; PIDN:CA098577.1; PID:gl6409874; GSPDB:GN00177  
 Experimental source: strain EGB-e  
 Genetics:  
 Gene: lmo0498  
 Superfamily: galactoside O-acetyltransferase

Query Match 100.0%; Score 19; DB 2; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 ||||  
 b 124 VAEF 127

## RESULT 72

AB2857  
 pseudoazurin [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 Species: Agrobacterium tumefaciens  
 Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 Accession: AB2857  
 Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.

Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 Reference number: AB2577; MUID:21608550; PMID:11743193

Y 1 VAEF 4  
 ||||  
 b 124 VAEF 127

AB2857  
 Status: preliminary  
 Molecule type: DNA  
 Residues: 1-150 <KUR>  
 Cross-references: GB:AB008698; PIDN:AAL43272.1; PID:gi7740759; GSPDB:GN00186  
 Experimental source: strain C58 (Dupont)  
 Genetics:

;Gene: Atu2283  
;Map position: circular chromosome  
;Superfamily: plastocyanin

Query Match 100.0%; Score 19; DB 2; Length 150;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
78 VAEF 81

35ULT 73

rotein gp35 from Bacteriophage A118 homolog lmo0113 [imported] - Listeria monocytogenes  
;Species: Listeria monocytogenes  
;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
;Accession: AB1089  
;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Psihi, H.  
; Jones, L.M.; Karst, U.  
; Science 294, 849-852, 2001  
; Authors: Kreft, J.; Kuhn, M.; Kunst, P.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
; C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
; Title: Comparative genomics of Listeria species.  
; Reference number: AB1077; MUID:21537279; PMID:11679669

;Accession: AB1089

;Status: preliminary

;Molecule type: DNA

;Residues: 1-150 <GLA>

;Cross-references: GB:NC\_003210; PIDN:CA98328.1; PID:gl6409472; GSPDB:GN00177

;Experimental source: strain EGD-e

;Genetics:

;Gene: lmo0113

Query Match 100.0%; Score 19; DB 2; Length 150;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
105 VAEF 108

RESULT 74

rotein gp35 from Bacteriophage A118 homolog lin0160 [imported] - Listeria innocua (stra  
;Species: Listeria innocua  
;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
;Accession: A11452  
;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Psihi, H.  
; Jones, L.M.; Karst, U.  
; Science 294, 849-852, 2001  
; Authors: Kreft, J.; Kuhn, M.; Kunst, P.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
; C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
; Title: Comparative Genomics of Listeria species.  
; Reference number: AB1077; MUID:21537279; PMID:11679669

;Accession: A11452

;Status: preliminary

;Molecule type: DNA

;Residues: 1-150 <GLA>

;Cross-references: GB:AL592022; PIDN:CA95393.1; PID:gl6412579; GSPDB:GN00178

;Experimental source: strain Clif11262

;Genetics:

;Gene: lin0160

Query Match 100.0%; Score 19; DB 2; Length 150;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4

Db 105 VAEF 108

RESULT 75

H87263  
hypothetical protein CC0121 [imported] - Caulobacter crescentus  
;Species: Caulobacter crescentus  
;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
;Accession: H87263

R.Nierman, W.C.; Reldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.;  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: H87263

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-151 <STO>

A:Cross-references: GB:AE005673; NID:gl3421228; PIDN:AAK22108.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0121

Query Match 100.0%; Score 19; DB 2; Length 151;

Best Local Similarity 100.0%; Pred. No. 4.3e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4

Db 73 VAEF 76

Search completed: May 24, 2004, 17:38:37

Job time : 11 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

in on: May 24, 2004, 17:33:28 ; Search time 6.85714 Seconds  
(without alignments)  
30.374 Million cell updates/sec

itle: US-09-594-978A-2

sequence: 1 VAEF 4

scoring table: BLOSUM62

sarched: Gapop 10.0 , Gapext 0.5

otal number of hits satisfying chosen parameters: 141681

imum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Listing first 1000 summaries

atabase : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	19	100.0	19	NS2 MYCTU	P81136 mycobacteri
2	19	100.0	59	SECE TREPA	O83263 treponema p
3	19	100.0	59	Y490 ARCFU	O29175 archaeoglob
4	19	100.0	63	Y431 HAEIN	P43998 haemophilus
5	19	100.0	88	RS15 HAEIN	P43398 haemophilus
6	19	100.0	93	ALB2 METKA	Q82we6 methanopyru
7	19	100.0	108	YF24 METJA	Q58919 methanococc
8	19	100.0	114	KEDA ACTSL	P41249 actinomycet
9	19	100.0	115	INS VERMO	Q9W7r2 verasper mo
10	19	100.0	116	INS LORPI	P01341 lophius pis
11	19	100.0	119	U279 DROME	Q9W8f3 drosophila
12	19	100.0	120	GTR2 PIG	Q62786 sus scrofa
13	19	100.0	122	R18E PYRAE	O82yq2 pyrobaculum
14	19	100.0	124	Y670 PASMU	Q9cmY0 pasteurella
15	19	100.0	127	ILBP MOUSE	P51162 mus musculu
16	19	100.0	130	SPFH THEMA	Q9wzc3 thermotoga
17	19	100.0	131	Y194 PYRAE	Q9v280 pyrococcus
18	19	100.0	133	R18E HALNI	Q9hrm5 halobacteri
19	19	100.0	134	YBGC ECOLI	P08999 escherichia
20	19	100.0	139	R18B METLI	O27443 methanobact
21	19	100.0	140	Y586 ARCFU	O29669 archaeoglob
22	19	100.0	143	R18E ARCFU	O28152 archaeoglob
23	19	100.0	145	RL13 HALMA	P29198 haloarcula
24	19	100.0	148	CLM4 MOUSE	Q9im83 mus musculu
25	19	100.0	148	YORX PYRMO	P20298 pyrococcus
26	19	100.0	150	R18B PYRAE	Q8zre3 pyrobaculum
27	19	100.0	152	SODC PRIGL	P14148 prionace g1
28	19	100.0	154	R18B SULSO	Q980b5 sulfolobus
29	19	100.0	155	PCP HAEIN	P10325 haemophilus
30	19	100.0	157	RAP TAROF	Q49065 taraxacum o
31	19	100.0	157	R18E SULTO	Q975m5 sulfolobus
32	19	100.0	163	SGSS DROME	P07701 drosophila
33	19	100.0	166	UTR5 YEAST	P32630 saccharomyc

34	19	100.0	169	SSB_RHOBA	P59932 rhodopirell
35	19	100.0	172	BFL1 MOUSE	Q07440 mus musculu
36	19	100.0	175	BFL1 HUMAN	Q16548 homo sapien
37	19	100.0	176	FR12_RANCA	P07798 rana catesb
38	19	100.0	177	RL10_THETN	Q8r7u4 thermoanaer
39	19	100.0	178	DSBB_PASMU	Q916b3 pasteurella
40	19	100.0	178	ESM5_DROME	P13096 drosophila
41	19	100.0	178	YB18_ARCFU	O29147 archaeoglob
42	19	100.0	179	APT HAEDU	Q7vkq4 haemophilus
43	19	100.0	183	RETB_BOVIN	P18902 bos taurus
44	19	100.0	186	DNAA_WOLSP	P35907 wolbachia s
45	19	100.0	190	UCR1_PARDE	P05417 paracoccus
46	19	100.0	193	NTPA_METJA	Q57679 methanococc
47	19	100.0	194	HAM2_BACHD	Q3k849 bacillus ha
48	19	100.0	194	TNR0_ECOLI	P05823 escherichia
49	19	100.0	196	PINR_ECOLI	P77170 escherichia
50	19	100.0	196	PINR_ECOLI	P77574 escherichia
51	19	100.0	196	UHPA_ECOLI	P10940 escherichia
52	19	100.0	196	UHPA_SALTU	P27667 salmonella
53	19	100.0	198	AICD_MOUSE	Q9wve0 mus musculu
54	19	100.0	199	HIS5_SULSO	Q33777 sulfolobus
55	19	100.0	201	RETB_HORSE	Q28369 equus cabal
56	19	100.0	201	RETB_HUMAN	P02753 homo sapien
57	19	100.0	201	RETB_PIG	P27485 sus scrofa
58	19	100.0	201	RETB_RABIT	P06912 oryctolagus
59	19	100.0	204	VIGZ_ECOLI	P27862 escherichia
60	19	100.0	205	MD21_HUMAN	Q13257 homo sapien
61	19	100.0	205	MD21_MOUSE	Q32185 mus musculu
62	19	100.0	208	YAC1_LEGPN	P37033 legionella
63	19	100.0	215	HMUO_CORDI	P71119 corynebacte
64	19	100.0	218	COAT_CMVIX	Q65120 cucumber mo
65	19	100.0	220	PTP1_NPVOP	O10274 orygia pseu
66	19	100.0	223	ECR1_ARCFU	O29758 archaeoglob
67	19	100.0	223	RAN_TETPY	P41914 tetrahymena
68	19	100.0	223	YAC1_LEGPN	Q37880 thermus the
69	19	100.0	225	RAN_TETTH	P41935 tetrahymena
70	19	100.0	228	Y132_NPVOP	P24080 orygia pseu
71	19	100.0	230	NADD_PSEPK	Q88d15 pseudomonas
72	19	100.0	230	RR3_FORPU	P51308 porphyra pu
73	19	100.0	230	URIE_COREF	Q8fab3 corynebacte
74	19	100.0	230	URIE_COREL	Q8nt39 corynebacte
75	19	100.0	236	YGIB_WOLSU	P24195 escherichia
76	19	100.0	236	PUR7_WOLSU	Q7m9X5 wolliella s
77	19	100.0	236	YF58_MYCTU	Q50740 mycobacteri
78	19	100.0	238	Y437_STAEP	Q8ctk9 staphylococ
79	19	100.0	238	Y669_STAAM	Q8nrx2 staphylococ
80	19	100.0	238	YH89_STRMU	Q8dsj4 streptococc
81	19	100.0	239	RLPA_NEIMA	Q9ism7 neisseria m
82	19	100.0	239	RLPA_NEIME	Q9k1a0 neisseria m
83	19	100.0	239	VG88_EFMLS	Q5305 mycobacteri
84	19	100.0	240	ATP7_ARATH	Q9sj12 arabidopsis
85	19	100.0	240	KOKA_PASPI	P58551 pasteurella
86	19	100.0	241	PCVA_PROMP	Q93tl5 prochloroco
87	19	100.0	242	Y755_AQUAE	Q66957 aquifex aeo
88	19	100.0	244	COMB_SYNEL	Q8dl5 synechococc
89	19	100.0	247	ATPD_SORBI	Q07300 sargham bic
90	19	100.0	247	CYSZ_VIBPA	Q87rj6 vibrio para
91	19	100.0	249	AQP_AEDAS	Q9rnw7 aedes aegypt
92	19	100.0	250	TIP1_TOBAC	P21653 nicotiana t
93	19	100.0	250	TIP2_TOBAC	P24422 nicotiana t
94	19	100.0	250	TIP2_ANTMA	P33560 antirrhinum
95	19	100.0	253	KUUD_ECOLI	P37769 escherichia
96	19	100.0	253	KUUD_ERWCH	Q05528 erwinia chr
97	19	100.0	255	PPI_BRANA	P23777 brassica na
98	19	100.0	255	TRME_PASMU	Q9clc2 pasteurella
99	19	100.0	257	PSD8_HUMAN	P48556 homo sapien
100	19	100.0	257	PSD8_MOUSE	Q9cx56 mus musculu
101	19	100.0	258	NADG_ARCFU	O28439 archaeoglob
102	19	100.0	261	STX7_HUMAN	O15400 homo sapien
103	19	100.0	261	STX7_MOUSE	O70439 mus musculu
104	19	100.0	261	STX7_RAT	O70257 rattus norv
105	19	100.0	266	TRPA_SYNEL	Q8dlm9 synechococc
106	19	100.0	267	APHC_MOUSE	Q9d099 mus musculu

107	19	100.0	267	1	1	DKGB_YERPE	Q82b36 versinia pe	180	19	100.0	325	1	UCPS_MOUSE	Q92b22 mus musculu
108	19	100.0	267	1	1	GNPI_CABEL	Q9xv12 caenorhabdi	181	19	100.0	326	1	PE46_ARATH	O81772 arabidopsis
109	19	100.0	267	1	1	MM07_RAT	P50280 rattus norv	182	19	100.0	326	1	YP87_CABEL	Q09444 caenorhabdi
110	19	100.0	269	1	1	AQPI1_HUMAN	P29972 homo sapien	183	19	100.0	328	1	AIPL_RAT	Q091g9 rattus norv
111	19	100.0	269	1	1	AQPI1_MOUSE	Q02013 mus musculu	184	19	100.0	328	1	Y778_AQUAE	O66974 aquifex aeo
112	19	100.0	269	1	1	RM09_YEAST	P29975 rattus norv	185	19	100.0	329	1	YFV6_SCHPO	Q9p7q8 schizosacch
113	19	100.0	269	1	1	SRB1_MOUSE	P311334 saccharomyc	186	19	100.0	330	1	PRC2_PORGI	Q58250 methanococc
114	19	100.0	269	1	1	YS13_STAAU	P47758 mus musculu	187	19	100.0	330	1	Y840_METJA	P25144 bacillus su
115	19	100.0	269	1	1	Q95337_MOUSE	Q05337 staphylococ	188	19	100.0	334	1	CCPA_BACSU	Q52544 rhizobium s
116	19	100.0	271	1	1	AQPI1_BOVIN	P47865 bos taurus	189	19	100.0	334	1	PRC1_PORGI	P59916 porphyromon
117	19	100.0	271	1	1	SRBP_HUMAN	Q9y5m8 homo sapien	190	19	100.0	334	1	YJH3_SHIME	Q92v44 rhizobium m
118	19	100.0	272	1	1	AQPI1_SHEEP	P56401 ovis aries	191	19	100.0	337	1	OPCM_CHICK	Q98892 gallus gall
119	19	100.0	272	1	1	YF63_STEAM	Q99tt3 staphylococ	192	19	100.0	338	1	Y4XL_RHISN	P58704 rhizobium s
120	19	100.0	274	1	1	AMYC_THETU	P37729 thermomanaer	193	19	100.0	339	1	HRDC_STRCO	P18184 streptomyce
121	19	100.0	277	1	1	Y087_THEMA	Q9wku2 thermotoga	194	19	100.0	339	1	IGH1_HUMAN	P78318 homo sapien
122	19	100.0	279	1	1	PHEA_LACLA	Q9ceu2 lactococcus	195	19	100.0	340	1	HUR_STRAU	Q61249 mus musculu
123	19	100.0	279	1	1	PHEA_LACLC	P43309 lactococcus	196	19	100.0	340	1	IGBI_MOUSE	O08836 rattus norv
124	19	100.0	282	1	1	TRUB_BORBU	O511743 borellia bu	197	19	100.0	340	1	IGBI_RAT	Q98836 rattus norv
125	19	100.0	284	1	1	Y3G4_SHEON	Q8ae22 shewanella	198	19	100.0	343	1	MO2M_ARATH	Q9m0m4 arabidopsis
126	19	100.0	284	1	1	YN13_MYCTU	P71899 mycobacteri	199	19	100.0	343	1	MO2M_ARATH	Q9f6k3 arabidopsis
127	19	100.0	285	1	1	YB46_HASIN	P45071 haemophilus	200	19	100.0	343	1	XT30_ARATH	Q38908 arabidopsis
128	19	100.0	287	1	1	PDXY_ECOLI	P77150 escherichia	201	19	100.0	344	1	CD2_MOUSE	P08921 rattus norv
129	19	100.0	287	1	1	SVQ_ECOLI	P14325 dictyosteli	202	19	100.0	344	1	CD2_RAT	P08921 rattus norv
130	19	100.0	288	1	1	Y169_PASMU	Q9CP85 pasteurella	203	19	100.0	344	1	STSY_RAUSE	P15324 raoultia s
131	19	100.0	291	1	1	AMPM_METJA	O28438 archaeoglob	204	19	100.0	345	1	TRPD_SULSO	P50384 sulfolobus
132	19	100.0	292	1	1	YG29_METJA	O59023 methanococc	205	19	100.0	347	1	FLIG_AGRTE	Q44588 agrobacteri
133	19	100.0	294	1	1	AMPM_METJA	O58725 methanococc	206	19	100.0	352	1	CYCR_ROSDE	P26278 roseobacte
134	19	100.0	294	1	1	SSB1_YEAST	P10080 saccharomyc	207	19	100.0	352	1	STSY_CATRO	P18417 catharanthu
135	19	100.0	295	1	1	PRMA_XANCP	Q9kv64 vibrio chol	208	19	100.0	354	1	DCDI_HUMAN	P58994 homo sapien
136	19	100.0	295	1	1	REBA_XANCP	P55256 xanthomonas	209	19	100.0	362	1	MURG_SHEON	Q8KX35 shewanella
137	19	100.0	295	1	1	YDH2_XANAU	P22444 xanthobacte	210	19	100.0	364	1	DDL_CHLTE	Q8K368 chlorobium
138	19	100.0	297	1	1	YD67_METJA	Q58762 methanococc	211	19	100.0	364	1	MRP_AQUAE	Q53560 bacillus li
139	19	100.0	299	1	1	T2R1_HUMAN	Q9nyw7 homo sapien	212	19	100.0	364	1	MRP_AQUAE	O66946 aquifex aeo
140	19	100.0	300	1	1	NARE_CHICK	O29080 gallus gall	213	19	100.0	366	1	CYCR_RHOGE	Q9f1n0 shewanella
141	19	100.0	304	1	1	CYSD_XYLFA	Q9p879 xylella fas	214	19	100.0	369	1	RPSP_STRPN	P43726 haemophilus
142	19	100.0	304	1	1	CYSD_XYLFA	Q87d98 xylella fas	215	19	100.0	369	1	KLAB_ECOLI	P43726 haemophilus
143	19	100.0	305	1	1	E2BA_CABEL	P34604 caenorhabdi	216	19	100.0	378	1	KLAB_ECOLI	P43726 haemophilus
144	19	100.0	308	1	1	CP3_ENTHI	Q06964 entamoeba h	217	19	100.0	379	1	TGT_VIBCH	Q9kt99 vibrio chol
145	19	100.0	308	1	1	XSGA_BIFLO	Q8g613 d dimorchyia	218	19	100.0	380	1	METE_HELPU	Q9zmv7 helicobacte
146	19	100.0	309	1	1	DNJL_MYCFN	O50312 mycoplasma	219	19	100.0	380	1	METE_HELPU	P56069 helicobacte
147	19	100.0	309	1	1	NTCB_SYNY3	P74422 synochocyst	220	19	100.0	381	1	CDBA_SCHPO	Q09184 schizosacch
148	19	100.0	310	1	1	MIAA_CLOPE	O8x185 clostridium	221	19	100.0	383	1	CYCR_CHRVI	O82947 chromatiu
149	19	100.0	310	1	1	PI12_TOBAC	O04857 nicotiana t	222	19	100.0	384	1	AIPL_HUMAN	Q9nm99 homo sapien
150	19	100.0	310	1	1	SIR5_HUMAN	Q9nxa8 homo sapien	223	19	100.0	384	1	AIPL_PANPA	Q95nm9 pan paniscu
151	19	100.0	310	1	1	SIR5_MOUSE	Q8k266 mus musculu	224	19	100.0	385	1	GBA5_CABEL	Q20701 caenorhabdi
152	19	100.0	310	1	1	Y942_SULSO	Q97zh1 sulfolobus	225	19	100.0	385	1	PGK_HASIN	P43726 haemophilus
153	19	100.0	311	1	1	DAPA_STRPN	Q97xr23 streptococc	226	19	100.0	387	1	VPD_BPP2	P10312 bacterioph
154	19	100.0	311	1	1	DAPA_STRRN6	Q8dpz3 streptococc	227	19	100.0	388	1	METE_MYCLE	P46807 mycobacteri
155	19	100.0	311	1	1	MIAA_LACPL	Q8dpz3 streptococc	228	19	100.0	388	1	METE_MYCTU	Q53427 mycobacteri
156	19	100.0	311	1	1	OSL1_HUMAN	Q88wp5 lactobacill	229	19	100.0	389	1	RF1M_CABEL	O44568 caenorhabdi
157	19	100.0	311	1	1	OSL2_HUMAN	Q8ng12 homo sapien	230	19	100.0	389	1	VPD_BP186	P21679 bacterioph
158	19	100.0	312	1	1	OSN1_HUMAN	Q8ng10 homo sapien	231	19	100.0	390	1	Y181_ARCFU	O30056 archaeoglob
159	19	100.0	312	1	1	PP12_ARATH	Q8ngv5 homo sapien	232	19	100.0	392	1	TEBB_STMY	P29548 stylonychia
160	19	100.0	312	1	1	PP15_ARATH	P48482 arabidopsis	233	19	100.0	392	1	YB1F_HASIN	P44083 haemophilus
161	19	100.0	313	1	1	DRPE_CRAPL	P22242 craterostig	234	19	100.0	392	1	YB1F_HASIN	Q55663 synochocyst
162	19	100.0	313	1	1	PEPM_STRHY	P29247 streptomyce	235	19	100.0	395	1	DXR_SVNY3	Q55663 synochocyst
163	19	100.0	314	1	1	IUNH_CRIFA	Q97546 crithidia f	236	19	100.0	397	1	CGA3_CHICK	P43449 gallus gall
164	19	100.0	314	1	1	PP1_EPAUL	P48490 phaseolus v	237	19	100.0	398	1	Y929_THEMA	P56727 thermotoga
165	19	100.0	315	1	1	OSAZ_HUMAN	P47893 homo sapien	238	19	100.0	399	1	PNCB_BUCAP	Q8K516 buchnera ap
166	19	100.0	315	1	1	OSAZ_PANTR	Q9cu37 pan troglod	239	19	100.0	400	1	ER19_HUMAN	P57442 buchnera ap
167	19	100.0	315	1	1	OSAZ_HUMAN	P47888 homo sapien	240	19	100.0	401	1	ER19_MOUSE	P53602 homo sapien
168	19	100.0	315	1	1	OSAZ_PANTR	Q9cu40 pan troglod	241	19	100.0	401	1	ER19_MOUSE	Q991f5 mus musculu
169	19	100.0	316	1	1	APB1_MOUSE	P28352 mus musculu	242	19	100.0	402	1	CL23_MYCTU	Q62967 rattus norv
170	19	100.0	316	1	1	APB1_MOUSE	P43138 rattus norv	243	19	100.0	402	1	CL13B_XYLFA	P77302 mycobacteri
171	19	100.0	316	1	1	APB1_MOUSE	P43138 rattus norv	244	19	100.0	402	1	CL13B_XYLFA	Q9p3x5 xylella fas
172	19	100.0	316	1	1	APB1_MOUSE	P48487 brassica ol	245	19	100.0	406	1	CGA2_BOVIN	P30274 bos taurus
173	19	100.0	317	1	1	APB1_MOUSE	P27695 homo sapien	246	19	100.0	406	1	T230_HUMAN	P48775 homo sapien
174	19	100.0	317	1	1	OSTP_RAT	Q08721 rattus norv	247	19	100.0	407	1	RPF2_HUMAN	O60858 homo sapien
175	19	100.0	317	1	1	OSTP_RAT	Q08721 rattus norv	248	19	100.0	409	1	Y4DM_RHISN	P55412 rhizobium s
176	19	100.0	318	1	1	PP11_TOBAC	O4856 nicotiana c	249	19	100.0	410	1	TDPI_HUMAN	Q14196 homo sapien
177	19	100.0	322	1	1	PP16_ARATH	P30366 arabidopsis	250	19	100.0	410	1	TDPI_MOUSE	Q08639 mus musculu
178	19	100.0	324	1	1	ADD_RHIME	P48486 arabidopsis	251	19	100.0	411	1	PEPT_YERPE	Q8zf10 versinia pe
179	19	100.0	324	1	1	ADD_RHIME	Q92t48 rhizobium m	252	19	100.0	412	1	CCA_ECOLI	P06961 escherichia
			325	1	1	UCPS_MOUSE	Q92b22 mus musculu							



399	19	100.0	580	1	RS1_CHLPP	Q928m3 chlamydia p	472	19	100.0	649	1	INVA_PHAU	P29001 phaseolus a
400	19	100.0	587	1	COBE_CNQW	Q90x85 oncorhynchu	473	19	100.0	651	1	INVA_PHAU	Q24509 phaseolus v
401	19	100.0	588	1	HAS1_XENUS	Q13563 xenopus lae	474	19	100.0	652	1	CBB_BACTU	Q06117 bacillus th
402	19	100.0	589	1	GP11_MOUSE	Q01514 mus musculu	475	19	100.0	660	1	FDL_DROME	Q8w8f3 drosophila
403	19	100.0	589	1	GP22_MOUSE	Q92066 mus musculu	476	19	100.0	660	1	TP6B_METJA	Q84344 methanococc
404	19	100.0	589	1	GP22_MOUSE	Q63663 rattus norv	477	19	100.0	676	1	NICA_ARATH	Q8gum5 arabidopsis
405	19	100.0	591	1	GP22_MOUSE	Q63663 rattus norv	478	19	100.0	677	1	NCPR_MOUSE	P37040 mus musculu
406	19	100.0	591	1	GP22_MOUSE	Q63663 rattus norv	479	19	100.0	677	1	NCPR_MOUSE	P00388 rattus norv
407	19	100.0	592	1	ALAM_YEAST	P52893 saccharomyc	480	19	100.0	680	1	DCP_SALTY	P27236 salmonella
408	19	100.0	602	1	PEX5_HUMAN	P50542 homo sapien	481	19	100.0	685	1	CSD_MYCPA	Q9k116 mycobacteri
409	19	100.0	602	1	TP6B_ARCFU	Q029605 archaeoglob	482	19	100.0	690	1	TRF2_SALSA	P80426 salmo salar
410	19	100.0	604	1	DNAB_BACNE	P05646 bacillus me	483	19	100.0	691	1	HEXA_ANGA	P80429 salmo salar
411	19	100.0	605	1	APM2_YEAST	P38700 saccharomyc	484	19	100.0	692	1	RECC_PASMU	Q17020 anopheles g
412	19	100.0	607	1	BKL3_MOUSE	Q8bh14 mus musculu	485	19	100.0	693	1	TRFE_RABIT	Q9cmb4 pasteurella
413	19	100.0	607	1	DNAB_LACIA	P42368 lactococcu	486	19	100.0	695	1	TRFE_RABIT	P19134 cryptotagus
414	19	100.0	607	1	DNAB_STRP3	Q8k624 streptococc	487	19	100.0	696	1	TRFE_RABIT	P09571 sus scrofa
415	19	100.0	607	1	DNAB_STRP3	Q8k624 streptococc	488	19	100.0	704	1	TRFE_HUMAN	P02787 homo sapien
416	19	100.0	607	1	DNAB_STRP3	Q8k624 streptococc	489	19	100.0	704	1	ICA_PIG	Q29545 sus scrofa
417	19	100.0	607	1	DNAB_STRP3	Q8k624 streptococc	490	19	100.0	705	1	TRFE_BOVIN	Q29443 bos taurus
418	19	100.0	608	1	BKL3_HUMAN	P95831 streptococc	491	19	100.0	705	1	EPG_BACTN	Q8a474 bacteroides
419	19	100.0	609	1	DNAB_ENTPA	Q835r7 enterococcu	492	19	100.0	705	1	PAL2_LITER	Q49836 lithospermu
420	19	100.0	610	1	DNAB_BACSH	Q69268 bacillus sp	493	19	100.0	706	1	DREB_MOUSE	Q8g886 mus musculu
421	19	100.0	610	1	DNAB_STAM	Q99tr7 staphylococ	494	19	100.0	706	1	DREB_MOUSE	Q8f983 leptospira
422	19	100.0	610	1	DNAB_STAM	P45554 staphylococ	495	19	100.0	707	1	TRFE_HORSE	P27425 equus caball
423	19	100.0	611	1	DNAB_BACAA	Q81182 bacillus an	496	19	100.0	707	1	DREB_RAT	Q07366 rattus norv
424	19	100.0	611	1	DNAB_BACR	Q81889 bacillus ce	497	19	100.0	707	1	PAL1_IPOBA	P14166 ipomoea bat
425	19	100.0	612	1	DNAB_LISNO	Q92bns listeria in	498	19	100.0	708	1	PAL1_IPOBA	Q23865 daucus caro
426	19	100.0	612	1	DNAB_LISNO	Q92bns listeria in	499	19	100.0	708	1	PAL2_IPOBA	Q42858 ipomoea bat
427	19	100.0	612	1	ILVD_XANAC	Q8pq10 xanthomonas	500	19	100.0	710	1	PAL1_LITER	P03070 simian viru
428	19	100.0	612	1	ILVD_XANAC	Q8pq10 xanthomonas	501	19	100.0	712	1	PAL2_TOBAC	Q49835 lithospermu
429	19	100.0	612	1	YCB1_MYCTU	Q11040 mycobacteri	502	19	100.0	712	1	PAL3_TOBAC	P35513 nicotiana t
430	19	100.0	614	1	DNAB_LACSN	Q8kml6 lactobacilli	503	19	100.0	714	1	ZW10_DROGR	P45733 nicotiana t
431	19	100.0	614	1	SPKE_SYNY3	P73515 synecocyst	504	19	100.0	715	1	FAOB_PSEFR	Q44219 drosophila
432	19	100.0	621	1	TP6B_METAC	Q8tqf7 methanosarc	505	19	100.0	715	1	SYGB_ANASP	P28793 p fatty oxi
433	19	100.0	621	1	TP6B_METAC	Q8tqf7 methanosarc	506	19	100.0	715	1	SYGB_ANASP	Q8vbt0 anabaena sp
434	19	100.0	622	1	C2AC_BACTU	Q45743 bacillus th	507	19	100.0	721	1	GLGK_MYCTU	Q10767 mycobacteri
435	19	100.0	624	1	COOS_METUA	Q58138 methanococc	508	19	100.0	722	1	SYGB_SYNY3	Q55890 synecocyst
436	19	100.0	624	1	P2L2_HABIN	Q9zn66 homo sapien	509	19	100.0	725	1	PAL1_ARATH	P35510 arabidopsis
437	19	100.0	625	1	TP22_HABIN	P44971 haemophilus	510	19	100.0	725	1	YFEA_ECOLI	P03690 bacterioph
438	19	100.0	627	1	CACP_CANTR	Q00614 candida tro	511	19	100.0	729	1	PURL_CHLTE	P23842 escherichia
439	19	100.0	627	1	CACP_COLLI	P52846 columbia liv	512	19	100.0	759	1	PURL_CHLTE	Q8kd17 chlorobium
440	19	100.0	631	1	NTP1_VACCA	Q57214 vaccinia vi	513	19	100.0	763	1	PSRA_WOLSU	P31075 wulinella s
441	19	100.0	631	1	NTP1_VACCC	P20637 vaccinia vi	514	19	100.0	764	1	ACOC_CUCMC	Q42669 cucumis mel
442	19	100.0	631	1	NTP1_VACCV	P05807 vaccinia vi	515	19	100.0	781	1	DPOL_ARCFU	Q29753 archaeoglob
443	19	100.0	631	1	NTP1_VARY	P33066 variola viru	516	19	100.0	787	1	OXAA_CHLTR	Q9pkx3 chlamydia m
444	19	100.0	632	1	NTP1_VARY	Q9814 myxoma viru	517	19	100.0	787	1	OXAA_CHLTR	Q10313 schizosacch
445	19	100.0	632	1	NTP1_SFVXA	Q9822 Shope fibro	518	19	100.0	804	1	TP6B_HALN1	Q9hr31 halobacteri
446	19	100.0	634	1	DNAB_HABIN	P43736 haemophilus	519	19	100.0	807	1	ANCA_MOUSE	Q91w96 mus musculu
447	19	100.0	634	1	NTP1_MCV1	Q98267 molluscum c	520	19	100.0	808	1	ANCA_MOUSE	Q91w96 mus musculu
448	19	100.0	635	1	DNAB_VIBCH	Q34241 vibrio chol	521	19	100.0	808	1	ANCA_MOUSE	Q91w96 mus musculu
449	19	100.0	635	1	DNAB_VIBCH	Q34241 vibrio chol	522	19	100.0	808	1	ANCA_MOUSE	Q91w96 mus musculu
450	19	100.0	636	1	DNAB_VIBVU	Q9vz80 anabaena sp	523	19	100.0	818	1	MUTS_THETH	Q56239 thermus the
451	19	100.0	636	1	DNAB_VIBVU	Q9vz80 anabaena sp	524	19	100.0	821	1	CAN3_HUMAN	Q8dl37 synecococc
452	19	100.0	636	1	MYB_MOUSE	Q7m85 vibrio vuln	525	19	100.0	821	1	SYFB_SYNEL	P14346 herpesviru
453	19	100.0	637	1	DNAB_BRUOV	P66876 mus musculu	526	19	100.0	835	1	UL52_HSVSA	P45353 pichia past
454	19	100.0	637	1	DNAB_BRUOV	Q5981 bruceella ov	527	19	100.0	842	1	HIS2_PICPA	P53532 corynebact
455	19	100.0	637	1	DNAB_BRUSU	Q8fxx2 bruceella su	528	19	100.0	852	1	CLPB_COSGL	P20594 rhizopus ol
456	19	100.0	637	1	DNAB_ECOLI	P04475 escherichia	529	19	100.0	858	1	CHSL_RHOL	P34552 caenorhabdi
457	19	100.0	637	1	DNAB_PSEAB	Q9hv43 pseudomonas	530	19	100.0	861	1	MUTS_BORBU	Q51737 bruceella bu
458	19	100.0	637	1	DNAB_SALTY	Q8z9r1 salmonella	531	19	100.0	862	1	MUTS_BORBU	Q19954 caenorhabdi
459	19	100.0	637	1	DNAB_SALTY	Q8z9r1 salmonella	532	19	100.0	876	1	VP41_CABEL	Q9cd81 lactococcu
460	19	100.0	637	1	DNAB_VIBFA	Q56073 salmonella	533	19	100.0	877	1	DP01_LACLA	Q32801 lactococcu
461	19	100.0	637	1	DNAB_VIBFA	Q56073 salmonella	534	19	100.0	877	1	DP01_LACLA	P59199 streptococc
462	19	100.0	638	1	DNAB_BUCBP	Q97rx3 vibrio para	535	19	100.0	877	1	DP01_STRP6	P59200 streptococc
463	19	100.0	638	1	DNAB_RHILE	Q17121 vibrio proc	536	19	100.0	881	1	HELI_HSV2H	P28277 herpes simp
464	19	100.0	638	1	DNAB_RHILE	P59565 buchnera ap	537	19	100.0	882	1	SYA_SYNEL	Q8dh56 synecococc
465	19	100.0	638	1	DNAB_SHIFL	Q33528 rhizobium l	538	19	100.0	889	1	SEX13_HUMAN	Q9un79 homo sapien
466	19	100.0	640	1	MYB_BOVIN	Q83mh5 shigella fl	539	19	100.0	894	1	SEC3_MOUSE	Q9nv70 homo sapien
467	19	100.0	640	1	MYB_BOVIN	Q87384 vibrio harv	540	19	100.0	894	1	SEC3_MOUSE	Q8r386 mus musculu
468	19	100.0	642	1	INVA_VICFA	P46200 bos taurus	541	19	100.0	895	1	PMAL_CANAL	P28877 candida alb
469	19	100.0	644	1	HS70_BRUMA	P20242 homo sapien	542	19	100.0	899	1	YKJ3_YEAST	Q04500 saccharomyc
470	19	100.0	644	1	LEU2_RHISA	Q43857 vicia faba	543	19	100.0	916	1	PMAL_AJESA	Q07421 ajellomyces
471	19	100.0	649	1	DREB_HUMAN	P17279 rhizomucor	544	19	100.0				
						Q16643 homo sapien							

545	19	100.0	916	1	SVI STAEAP	Q8csx1 staphylococ	618	19	100.0	2226	1	POLG_HPAV4	P26581 hepatitis a
546	19	100.0	919	1	Y893_HUMAN	O4967 homo sapien	619	19	100.0	2227	1	POLG_HPAV8	P26582 hepatitis a
547	19	100.0	920	1	DMR1_SCHPO	P78953 schizosacch	620	19	100.0	2228	1	POLG_HPAV8	P08617 hepatitis a
548	19	100.0	920	1	PRM1_NEUR	P07038 neurospora	621	19	100.0	2229	1	POLG_HPAVL	P04441 hepatitis a
549	19	100.0	921	1	SVI_BACSU	Q45477 bacillus su	622	19	100.0	2230	1	POLG_HPAVS	P14553 simian hepa
550	19	100.0	922	1	DPO1_RICFE	P87137 schizosacch	623	19	100.0	2231	1	PRF2_HUMAN	P23471 homo sapien
551	19	100.0	929	1	YDM6_SCHPO	O24617 arabidopsis	624	19	100.0	2236	1	PRF2_RAT	P62656 rattus norv
552	19	100.0	937	1	MSH2_ARATH	O27552 methanobact	625	19	100.0	2649	1	P285_HUMAN	Q95y88 homo sapien
553	19	100.0	937	1	SVL_METH	O27552 methanobact	626	19	100.0	2725	1	FLNC_HUMAN	Q14315 homo sapien
554	19	100.0	942	1	MSH2_MAIZE	Q9xgc9 zea mays (m	627	19	100.0	2774	1	MAPA_RAT	P34926 rattus norv
555	19	100.0	952	1	LYAG_HUMAN	P20253 homo sapien	628	19	100.0	2805	1	MAPA_HUMAN	P78559 homo sapien
556	19	100.0	956	1	SVY_VIBCH	Q9kr73 vibrio chol	629	19	100.0	3412	1	POLG_TBVS	P07720 t genome po
557	19	100.0	966	1	VIL3_ARATH	O81645 arabidopsis	630	19	100.0	3414	1	POLG_LANVT	P29837 t genome po
558	19	100.0	969	1	SVL_MYCTU	P71698 mycobacteri	631	19	100.0	3414	1	POLG_TBVEW	Q01239 t genome po
559	19	100.0	971	1	AMEN_HAECO	Q10737 haemochus	633	19	100.0	3415	1	POLG_POWVL	P14336 t genome po
560	19	100.0	972	1	MTFA_YEAST	Q98m36 rhizobium l	634	19	100.0	3803	1	TRAP_HUMAN	Q04538 t genome po
561	19	100.0	973	1	UVRA_PHILO	Q98m36 rhizobium l	635	19	100.0	3859	1	PKSK_BACSU	Q81u87 drosophila
562	19	100.0	974	1	POBJ_CAUCR	Q2z988 caulobacter	636	19	100.0	4447	1	MUC2_HUMAN	Q02817 homo sapien
563	19	100.0	976	1	VIL2_ARATH	O81644 arabidopsis	637	19	100.0	5179	1	MUC2_HUMAN	Q01896 cochiobolu
564	19	100.0	980	1	POLG_LIV	P22338 louping ill	638	19	100.0	5217	1	HIS1_HUMAN	Q14868 homo sapien
565	19	100.0	987	1	UVRA_XANCP	Q9pn26 xanthomonas	639	19	100.0	5262	1	MLL2_HUMAN	Q05055 tetrahymena
566	19	100.0	988	1	UVRA_XANCP	Q9pn26 xanthomonas	640	19	100.0	5596	1	MDN1_HUMAN	Q05055 tetrahymena
567	19	100.0	992	1	UVRA_MICLU	P13567 micrococcc	641	18	94.7	12	1	YOZD_BACSU	Q31863 bacillus su
568	19	100.0	997	1	CBF_MOUSE	P53569 mus musculu	642	18	94.7	58	1	YOZD_BACSU	Q8e6m0 streptococ
569	19	100.0	998	1	CBF_HUMAN	Q03701 homo sapien	643	18	94.7	71	1	EX7S_STRA3	Q8e6m0 streptococ
570	19	100.0	1012	1	DPOG_PICPA	Q01941 pichia past	644	18	94.7	79	1	DLTC_STPA3	Q8e6m0 streptococ
571	19	100.0	1021	1	MAPA_MOUSE	Q9gyr6 mus musculu	645	18	94.7	79	1	DLTC_STPA3	Q8e6m0 streptococ
572	19	100.0	1063	1	PDRI_YEAST	P12383 saccharomyc	646	18	94.7	81	1	YMG4_ECOLI	Q8e6m0 streptococ
573	19	100.0	1065	1	SMCS_SCHPO	O13710 schizosacch	647	18	94.7	88	1	YFCL_ECOLI	P41120 photorhabd
574	19	100.0	1086	1	RNC_CAREL	O01326 caenorhabdi	648	18	94.7	92	1	YFCL_ECOLI	P76496 escherichia
575	19	100.0	1095	1	AT9E_HUMAN	O43861 homo sapien	649	18	94.7	94	1	SELK_HUMAN	Q9y6d0 homo sapien
576	19	100.0	1095	1	AT9E_HUMAN	Q98195 mus musculu	650	18	94.7	94	1	SELK_MOUSE	Q911j1 mus musculu
577	19	100.0	1117	1	CYT4_NEUCR	P47950 neurospora	651	18	94.7	94	1	SELK_RAT	P59798 rattus norv
578	19	100.0	1123	1	V120_HSV1	P10221 herpes simp	652	18	94.7	95	1	R86_BACHD	Q9k5n8 bacillus ha
579	19	100.0	1125	1	MAPA_MOUSE	P27546 mus musculu	653	18	94.7	98	1	YCI1_ECOLI	P31070 escherichia
580	19	100.0	1144	1	FLNC_MOUSE	Q8vbx6 mus musculu	654	18	94.7	101	1	CYAY_HAETN	P71358 haemophilus
581	19	100.0	1146	1	YH3C_YEAST	P38472 saccharomyc	655	18	94.7	108	1	PRIB_NITEU	Q82xq7 nitrosomona
582	19	100.0	1155	1	IFP2_VETUA	Q57710 methanococ	656	18	94.7	110	1	VAG1_TOBAC	O82702 nicotiana t
583	19	100.0	1167	1	CAGA_HELPU	Q941t1 helicobacte	657	18	94.7	111	1	VAG2_TOBAC	O82703 nicotiana t
584	19	100.0	1173	1	UBAB_MOUSE	Q96800 mus musculu	658	18	94.7	117	1	RNPA_LACIA	Q9c173 lactococcus
585	19	100.0	1200	1	MOGS_CABEL	Q09530 caenorhabdi	659	18	94.7	133	1	RNPA_LACIA	Q9c173 lactococcus
586	19	100.0	1211	1	DP3A_HELPU	Q9zj19 helicobacte	660	18	94.7	134	1	YK49_STROO	P16251 streptomyce
587	19	100.0	1211	1	DP3A_HELPU	P56157 helicobacte	661	18	94.7	136	1	CDL_BACPY	Q9s3m0 bacillus ps
588	19	100.0	1239	1	TOP2_CRIFA	P27570 crithidia f	662	18	94.7	136	1	RL19_XYLFA	Q9ph36 xyella fas
589	19	100.0	1302	1	UBAB_HUMAN	O95155 homo sapien	663	18	94.7	136	1	RL19_XYLFT	Q87f53 xyella fas
590	19	100.0	1311	1	SPR4_CABEL	O17582 caenorhabdi	664	18	94.7	141	1	ALL2_TYRPU	Q02380 tyrophagus
591	19	100.0	1331	1	CYAB_LEIDO	Q25263 leishmania	665	18	94.7	141	1	R1SB_METUA	Q57751 methanococ
592	19	100.0	1355	1	DP3A_SYNY3	P74750 synechocyst	666	18	94.7	144	1	Y911_METUA	Q58398 methanococ
593	19	100.0	1371	1	VCAP_HSVTA	Q00999 herpesvirus	667	18	94.7	145	1	YK14_VIBVU	Q8db14 vibrio vuln
594	19	100.0	1423	1	ALBU_PETMA	Q91274 petromyzon	668	18	94.7	147	1	CALM_YEAST	O60041 kluyveromyc
595	19	100.0	1426	1	NPH4_HUMAN	O75161 homo sapien	669	18	94.7	147	1	CALM_YEAST	P06787 saccharomyc
596	19	100.0	1444	1	DPO3_LISIN	Q9c34 listeria in	670	18	94.7	147	1	YBSE_BACSU	P27163 petunia hyb
597	19	100.0	1444	1	DPO3_LISMO	Q8y7g1 listeria mo	671	18	94.7	148	1	CAL2_PETHY	O31511 bacillus su
598	19	100.0	1477	1	YORI_YEAST	P53049 saccharomyc	672	18	94.7	148	1	CAL2_PETHY	P49258 drosophila
599	19	100.0	1515	1	YCFI_YEAST	O88563 rattus norv	673	18	94.7	148	1	CALM_ACHKL	P15094 achlya kleb
600	19	100.0	1522	1	MRP3_RAT	Q14146 homo sapien	674	18	94.7	148	1	CALM_ACHKL	Q9hfy6 blasoccladi
601	19	100.0	1524	1	Y133_HUMAN	O15438 homo sapien	675	18	94.7	148	1	CALM_DROME	P07181 drosophila
602	19	100.0	1527	1	MRP3_HUMAN	O15823 caenorhabdi	676	18	94.7	148	1	CALM_DROME	P02594 electrophor
603	19	100.0	1535	1	LM11_CABEL	P30015 escherichia	677	18	94.7	148	1	CALM_ELEEL	P25996 metridium s
604	19	100.0	1538	1	LHR_ECOLI	O53114 mycobacteri	678	18	94.7	148	1	CALM_HORVU	P11118 euglena gra
605	19	100.0	1609	1	CTFI_MYCLE	Q8ch93 mus musculu	679	18	94.7	148	1	CALM_HORVU	P13565 hordeum vul
606	19	100.0	1679	1	GCC2_MOUSE	Q8ym40 anabaena sp	680	18	94.7	148	1	CALM_MAIZE	P02593 homo sapien
607	19	100.0	1906	1	YFAO_ANASP	Q62812 rattus norv	681	18	94.7	148	1	CALM_MAIZE	P27161 lycopersico
608	19	100.0	1961	1	MYH8_RAT	Q81zdg homo sapien	682	18	94.7	148	1	CALM_ORYSA	P07463 paramecium
609	19	100.0	2027	1	DOG3_MOUSE	Q8ciq7 mus musculu	683	18	94.7	148	1	CALM_ORYSA	P02595 patinopete
610	19	100.0	2030	1	DOG3_HUMAN	O81zdg homo sapien	684	18	94.7	148	1	CALM_PATSP	P27165 phytophor
611	19	100.0	2039	1	TEGU_HSV7U	Q9hzm1 homo sapien	685	18	94.7	148	1	CALM_PATSP	P11121 poridae sp
612	19	100.0	2061	1	MYOF_HUMAN	P28887 human respi	686	18	94.7	148	1	CALM_PATSP	P13868 solanum tub
613	19	100.0	2165	1	RRPL_HRSVA	Q69014 manchester	687	18	94.7	148	1	CALM_PATSP	P04353 spinacia oi
614	19	100.0	2208	1	POLN_MANCV	P20430 tacaribe vi	688	18	94.7	148	1	CALM_SPIOL	P21251 stichopos j
615	19	100.0	2210	1	RRPO_TACV	P21951 saccharomyc	689	18	94.7	148	1	CALM_SPIOL	
616	19	100.0	2222	1	DPOE_YEAST	P25580 hepatitis a	690	18	94.7	148	1	CALM_SPIOL	
617	19	100.0	2226	1	POLG_HPAV2								

691	18	94.7	148	1	CALM STYLE	P27166 stylonychia	764	18	94.7	229	1	RADP_THERAC	Q9hld3 thermoplasma
692	18	94.7	148	1	CALM TETPY	P02598 tetrahymena	765	18	94.7	230	1	ISPD_SINY3	P74223 synecocyst
693	18	94.7	148	1	CALM CHICK	P02597 gallus gall	766	18	94.7	231	1	FP_CKIMI	P15697 cricetus
694	18	94.7	149	1	CALM CANAL	P23286 candida alb	767	18	94.7	234	1	SAMP_MESAU	P07629 mesocricetu
695	18	94.7	149	1	CALM WHEAT	P04464 triticum ae	768	18	94.7	235	1	ICLN_CANFA	P35521 canis famil
696	18	94.7	149	1	MLE3_RABIT	P06741 homo sapien	769	18	94.7	236	1	PUR7_HELHP	Q7vfq6 helicobacte
697	18	94.7	149	1	MLE3_RABIT	P02603 oryctolagus	770	18	94.7	237	1	PUR7_METAC	Q8tis9 methanosarc
698	18	94.7	149	1	MLE3_RAT	P02601 rattus norv	771	18	94.7	237	1	PUR7_METMA	Q8yvk6 methanosarc
699	18	94.7	150	1	CALM SCHPO	P05933 schizosacch	772	18	94.7	238	1	RLI_ANASP	Q8y1j7 anabaena sp
700	18	94.7	150	1	CALM DICTDI	P02797 methanococc	773	18	94.7	238	1	Y53E_CHLTR	Q851j7 chlamydia t
701	18	94.7	151	1	CALM DICTDI	Q9nlp6 drosophila	774	18	94.7	238	1	Y53E_CHLTR	Q851j7 chlamydia t
702	18	94.7	151	1	CP2E_DROME	P23486 rhizobium m	775	18	94.7	240	1	Y82S_CHLMO	Q9PJk4 chlamydia m
703	18	94.7	151	1	Y4G6_RHIME	P50858 enterobacte	776	18	94.7	242	1	PUR7_COXBU	Q83ca8 coxiella bu
704	18	94.7	152	1	AAC6_ENTAE	P34440 naegleria g	777	18	94.7	242	1	MTGA_ECOLI	P46022 escherichia
705	18	94.7	155	1	CALF_NABGR	P02593 fusobacteri	778	18	94.7	242	1	MTGA_ECOLI	Q48465 klebsiella
706	18	94.7	155	1	Y463_FUSNR	P02593 fusobacteri	779	18	94.7	242	1	PDJX_THETM	Q8t9v9 thermosarc
707	18	94.7	159	1	TPCS_HUMAN	P02585 homo sapien	780	18	94.7	242	1	PDJX_THETM	Q8t9v9 thermosarc
708	18	94.7	159	1	TPCS_HUMAN	P02801 mus musculu	781	18	94.7	242	1	PDJX_THETM	Q8t9v9 thermosarc
709	18	94.7	159	1	TPCS_PIG	P02587 sus scrofa	782	18	94.7	244	1	Y82S_CHLMO	Q851j7 chlamydia t
710	18	94.7	159	1	TPCS_RABIT	P02586 oryctolagus	783	18	94.7	244	1	Y82S_CHLMO	Q851j7 chlamydia t
711	18	94.7	160	1	TPCS_ANGAN	P81660 anguilla an	784	18	94.7	245	1	YCDX_ECOLI	Q2B148 archaeoglob
712	18	94.7	162	1	AROK_LACLA	Q9ceul lactococcus	785	18	94.7	246	1	YCDX_ECOLI	P75914 escherichia
713	18	94.7	162	1	CALM CHICK	P04352 chlamydomon	786	18	94.7	246	1	AQPM_METH	O26206 methanobact
714	18	94.7	162	1	TPCS_CHICK	P02588 gallus gall	787	18	94.7	246	1	AQPM_METH	Q9c425 methanobact
715	18	94.7	162	1	TPCS_MELGA	P10246 meleagris g	788	18	94.7	246	1	HEM4_CHLVI	Q59335 chlorobium
716	18	94.7	162	1	TPCS_RANES	P02589 rana escul	789	18	94.7	246	1	MTGA_HAEIN	P44890 haemophilus
717	18	94.7	166	1	Y531_SULSO	Q8uw66 sulfolobus	790	18	94.7	248	1	TPRC_SULTO	Q972a1 sulfolobus
718	18	94.7	166	1	Y628_METJA	Q8uw66 sulfolobus	791	18	94.7	248	1	Y4EL_RHISN	P55435 rhizobium s
719	18	94.7	169	1	IPR3_ANASP	P08045 methanococ	792	18	94.7	251	1	TPIS_PSESY	Q8wq1 pseudomonas
720	18	94.7	169	1	IPR3_SINY3	P08057 synecocyst	793	18	94.7	252	1	TPIS_PSESY	P5576 pseudomonas
721	18	94.7	171	1	NSG2_HUMAN	Q9y328 homo sapien	794	18	94.7	252	1	RECO_CHLNP	Q9z7w5 chlamydia p
722	18	94.7	171	1	NSG2_MOUSE	P47759 mus musculu	795	18	94.7	253	1	TPRC_THEMA	Q56319 thermotoga
723	18	94.7	171	1	YB2_ACIAM	P32986 acidianus a	796	18	94.7	254	1	YTBQ_BACSU	P53560 bacillus su
724	18	94.7	172	1	IPYR_SYNEL	Q8dhx2 synecococ	797	18	94.7	254	1	AROD_NEIMB	Q91y60 neisseria m
725	18	94.7	173	1	TPCA_ABMME	P50157 ambystoma m	798	18	94.7	254	1	PWM_YEAST	Q91y60 neisseria m
726	18	94.7	173	1	MLRN_DROME	P40433 drosophila	799	18	94.7	257	1	RP38_BACAA	P07283 saccharomyc
727	18	94.7	177	1	IPYR_HALNI	Q9hs43 halobacteri	800	18	94.7	257	1	UXUR_ECOLI	Q9x5j6 bacillus an
728	18	94.7	181	1	PYR3_THEQP	P96078 thermus aqu	801	18	94.7	257	1	UXUR_ECOLI	P32161 escherichia
729	18	94.7	184	1	Y803_CHLNP	Q9z7a3 chlamydia p	802	18	94.7	260	1	EXUR_ERWCH	P42608 escherichia
730	18	94.7	185	1	CPI1_HUMAN	P23197 homo sapien	803	18	94.7	260	1	RS3_ANASP	Q9x9e0 erwinia chr
731	18	94.7	185	1	DX1_HUMAN	Q00765 homo sapien	804	18	94.7	261	1	TRPC_BACTN	Q8y1p5 anabaena sp
732	18	94.7	185	1	DPI_MOUSE	Q60870 mus musculu	805	18	94.7	261	1	RP38_BACSU	Q8aad6 bacteroides
733	18	94.7	185	1	NSG1_HUMAN	P42857 homo sapien	806	18	94.7	261	1	YNY6_YEAST	P06574 bacillus su
734	18	94.7	185	1	NSG1_MOUSE	Q62022 mus musculu	807	18	94.7	261	1	YNY6_YEAST	P53853 saccharomyc
735	18	94.7	185	1	RP3C_ECOLI	P37745 escherichia	808	18	94.7	261	1	UXUR_HAEIN	Q03768 saccharomyc
736	18	94.7	192	1	Y56A_THEMA	P87008 thermotoga	809	18	94.7	267	1	APHC_HUMAN	P44487 haemophilus
737	18	94.7	193	1	KTHY_THEAC	Q9h1z2 thermoplas	810	18	94.7	267	1	APHC_HUMAN	Q9xua6 sulfolobus
738	18	94.7	199	1	Y249_METTH	Q26351 methanobact	811	18	94.7	268	1	YG95_HAEIN	Q48215 haemophilus
739	18	94.7	200	1	ESPE_VIBMI	Q17172 vibrio mimi	812	18	94.7	268	1	HIS9_BACSU	O34411 bacillus su
740	18	94.7	201	1	RETB_MOUSE	Q07792 vibrio mimi	813	18	94.7	268	1	THIM_ARCFU	O28204 archaeoglob
741	18	94.7	201	1	RETB_MOUSE	Q07792 mus musculu	814	18	94.7	269	1	TRPA_BACST	O28204 archaeoglob
742	18	94.7	201	1	RETB_MOUSE	P04916 rattus norv	815	18	94.7	270	1	RPNC_SCHPO	P19867 bacillus st
743	18	94.7	202	1	COAB_CHLTR	Q94916 rattus norv	816	18	94.7	270	1	YCB7_METJA	P50524 schizosacch
744	18	94.7	202	1	HIS5_METAC	Q8t991 methanosarc	817	18	94.7	272	1	3DHO_ACICA	Q86683 methanococ
745	18	94.7	202	1	HIS5_METJA	Q9pvd5 methanosarc	818	18	94.7	272	1	AQPA_RANES	Q59087 acinetobact
746	18	94.7	202	1	Y555_MYCTU	Q11063 mycobacteri	819	18	94.7	275	1	SC65_CANAL	P50501 rana escul
747	18	94.7	206	1	CNC2_MOUSE	Q9db76 mus musculu	820	18	94.7	277	1	Y154_ARCFU	O14415 candida alb
748	18	94.7	208	1	RK1_HUMAN	Q9y3b6 homo sapien	821	18	94.7	277	1	Y154_ARCFU	O14415 candida alb
749	18	94.7	208	1	RK1_HUMAN	P49255 cavia porce	822	18	94.7	278	1	YD15_STAM	Q30083 archaeoglob
750	18	94.7	209	1	RE3_GRATE	P02743 homo sapien	823	18	94.7	278	1	YD15_STAM	Q30083 archaeoglob
751	18	94.7	216	1	TRPF_METJA	Q8t9x9 methanopyru	824	18	94.7	279	1	YD15_STAM	Q30083 archaeoglob
752	18	94.7	217	1	DEF1_BIFLO	Q8g534 bifidobacte	825	18	94.7	280	1	YD15_STAM	Q30083 archaeoglob
753	18	94.7	218	1	RECA_ARTAU	Q9re16 artrobacte	826	18	94.7	281	1	GLPF_ECOLI	P3004 arabidopsis
754	18	94.7	221	1	MLR_DROME	P18432 drosophila	827	18	94.7	281	1	GLPF_ECOLI	P1244 escherichia
755	18	94.7	223	1	DEOC_MYCCE	P47296 mycoplasma	828	18	94.7	282	1	PIP1_ATRCA	P31140 shigella fl
756	18	94.7	223	1	SAMP_HUMAN	P49255 cavia porce	829	18	94.7	283	1	LYSY_HAEIN	P42767 atriplex ca
757	18	94.7	223	1	SAMP_HUMAN	P49255 cavia porce	830	18	94.7	283	1	LYSY_HAEIN	Q2k420 haemophilus
758	18	94.7	224	1	DEOC_MYCPN	P02743 homo sapien	831	18	94.7	284	1	LYSY_PASMU	P44420 haemophilus
759	18	94.7	224	1	SAMP_PIG	Q19063 sus scrofa	832	18	94.7	284	1	LYSY_PASMU	P44420 haemophilus
760	18	94.7	226	1	ISGD_SALTY	P58654 salmonella	833	18	94.7	284	1	LYSY_PASMU	Q25609 helicobacte
761	18	94.7	228	1	ISGD_SALTY	Q8y1x9 anabaena sp	834	18	94.7	285	1	LYSY_PASMU	Q25609 helicobacte
762	18	94.7	228	1	IPYR_METTH	O26232 methanobact	835	18	94.7	286	1	LYSY_PASMU	Q25609 helicobacte
763	18	94.7	228	1	SAMP_RAT	P23680 rattus norv	836	18	94.7	286	1	LYSY_PASMU	Q25609 helicobacte

837	18	94.7	286	1	PI25 ARATH	Q9SV31 arabidopsis	910	18	94.7	363	1	LEU3_BUCAP	085064 buchnera ap
838	18	94.7	286	1	PI1 LYCES	Q08451 lycopersico	911	18	94.7	363	1	MURG_ENTFA	O07109 enterococcus
839	18	94.7	287	1	PI14 ARATH	Q39196 arabidopsis	912	18	94.7	364	1	AAT_PYRKO	Q93744 pyrococcus
840	18	94.7	287	1	PI15 ARATH	Q41a6 arabidopsis	913	18	94.7	364	1	YM28 MYCTU	O10512 mycobacteri
841	18	94.7	287	1	PI21 ARATH	Q3286 arabidopsis	914	18	94.7	365	1	HI82_BORBR	Q7wY3 bordetella
842	18	94.7	289	1	PI26 ARATH	Q42v07 arabidopsis	915	18	94.7	365	1	HI82_BORPA	Q7wY3 bordetella
843	18	94.7	289	1	PI22 DEA	P25794 pisum sativ	916	18	94.7	365	1	HI82_BORPE	Q7vS0 bordetella
844	18	94.7	291	1	EGC1_LISMO	Q9Y680 listeria mo	917	18	94.7	366	1	RECA_SPIPL	P48293 spirulina p
845	18	94.7	291	1	PI24 ARATH	Q9Y680 listeria mo	918	18	94.7	366	1	EDHA_BOVIN	P41563 bos taurus
846	18	94.7	294	1	ISPA_BUCAP	Q9Y680 listeria mo	919	18	94.7	366	1	EDHA_HUMAN	P50213 homo sapien
847	18	94.7	296	1	NI11 ARATH	Q8VZw1 arabidopsis	920	18	94.7	369	1	MEB_HABIN	P45022 haemophilus
848	18	94.7	296	1	YK23_ARCFU	Q8VZw1 arabidopsis	921	18	94.7	369	1	RECA_CORPS	P48288 corynebacte
849	18	94.7	298	1	PSD_CHLVC	Q82156 archaeoglob	922	18	94.7	369	1	VP6_AHSV3	P46909 african hor
850	18	94.7	300	1	Y025_NPVOP	Q82113 chlamydomo	923	18	94.7	369	1	VP6_AHSV6	Q64913 african hor
851	18	94.7	302	1	RPO4_FOWPV	O10298 oxygia pseu	924	18	94.7	371	1	YLB0_CLOAB	Q4354 closotridum
852	18	94.7	303	1	EST_AGLW	Q9Y540 fowlpox vir	925	18	94.7	372	1	BIOF_METJA	Q58694 methanococ
853	18	94.7	304	1	T2RD_HUMAN	P18773 acinetobact	926	18	94.7	372	1	DEGT_BACST	P15263 bacillus st
854	18	94.7	304	1	DCAS_AGRU	Q9YV9 homo sapien	927	18	94.7	376	1	RECA_CORGL	P24442 corynebacte
855	18	94.7	304	1	KHSE_STAAW	Q44185 agrobacteri	928	18	94.7	378	1	PHNM_ECOLI	P16689 escherichia
856	18	94.7	304	1	KHSE_STAAW	Q99ue6 staphylococ	929	18	94.7	380	1	METC_LACLA	Q9ra99 lactococcus
857	18	94.7	305	1	T2RD_RAT	Q8VW8 staphylococ	930	18	94.7	380	1	VH07_PASU	Q9ckb7 pasteurella
858	18	94.7	306	1	TRUB_NEIMA	Q9Jkt7 rattus norv	931	18	94.7	382	1	MATB_NEUCR	P36981 neurospora
859	18	94.7	307	1	TRUB_NEIMA	Q9Jkt5 neisseria m	932	18	94.7	386	1	RECA_COREF	Q9fpd3 corynebacte
860	18	94.7	310	1	DNUL_MYCSE	Q9JYV1 neisseria m	933	18	94.7	386	1	RNAL_SCHPO	P41391 schistosacch
861	18	94.7	310	1	YBR1_YEAST	P47248 mycoplasma	934	18	94.7	387	1	DNAT_CITFR	P45513 citrobacter
862	18	94.7	312	1	ARC2_ENTFA	P38238 saccharomyc	935	18	94.7	389	1	GAL1_FUSNN	Q8rhd0 fusobacteri
863	18	94.7	312	1	VI01_VACCC	P59625 enterococcu	936	18	94.7	390	1	Y4QE_RHISN	P55626 rhizobium s
864	18	94.7	312	1	VI01_VARV	P30498 vaccinia vi	937	18	94.7	393	1	CYS3_YEAST	P31373 saccharomyc
865	18	94.7	313	1	PEPM_STRVR	P22939 variola vir	938	18	94.7	393	1	HEMZ_YEAST	P16622 saccharomyc
866	18	94.7	313	1	Y376_AQUAE	O86937 streptomyc	939	18	94.7	394	1	AAT_AQUAE	O6781 aquifex aeo
867	18	94.7	315	1	O3A1_HUMAN	O66698 aquifex aeo	940	18	94.7	394	1	VORA_PYRPU	Q51801 pyrococcus
868	18	94.7	315	1	O3A1_PANTR	P47881 homo sapien	941	18	94.7	394	1	VORA_PYRHO	Q58413 pyrococcus
869	18	94.7	316	1	DHAS_VIBMI	Q9tuae pan troglod	942	18	94.7	395	1	VORA_PYRAB	Q9uyz1 pyrococcus
870	18	94.7	316	1	Y151_PYRTU	Q60080 vibrio mimi	943	18	94.7	397	1	PGK_STRPB	Q8uzg3 streptococ
871	18	94.7	318	1	Y211_AQUAE	Q8t2x4 pyrococcus	944	18	94.7	397	1	PGK_STRPV	P82487 streptococ
872	18	94.7	321	1	LPXD_BACSE	O86405 aquifex aeo	945	18	94.7	401	1	ASSV_AQUAE	O57213 aquifex aeo
873	18	94.7	322	1	YQYA_CAMCU	Q9phuo campylobact	946	18	94.7	404	1	HISX_ARCFU	O30027 archaeoglob
874	18	94.7	326	1	FSTL_ARATH	P54538 bacillus su	947	18	94.7	404	1	ISCS_EUCBP	Q9raa5 buchnera ap
875	18	94.7	330	1	LDHD_STR3	P52839 arabidopsis	948	18	94.7	405	1	ENTM_BACFR	P54355 bacteroides
876	18	94.7	330	1	LDHD_STR5	Q8e6a9 streptococ	949	18	94.7	405	1	Y872_METJA	Q58282 methanococ
877	18	94.7	330	1	LDHD_STRPY	Q8e0n5 streptococ	950	18	94.7	406	1	DEOB_AGRU5	Q8uJ04 agrobacteri
878	18	94.7	333	1	LDHD_LACPA	Q92m2 streptococ	951	18	94.7	406	1	DEOB_RHIME	Q92t47 rhizobium m
879	18	94.7	336	1	PEXG_HUMAN	P17584 lactobacill	952	18	94.7	408	1	YFIN_ECOLI	P46139 escherichia
880	18	94.7	336	1	RLI_MYCPE	Q9Y5Y5 homo sapien	953	18	94.7	408	1	YLO1_SCHPO	O13883 schizosacch
881	18	94.7	337	1	DHAS_VIBCH	Q8ex24 mycoplasma	954	18	94.7	410	1	YX28_MYCTU	Q50700 mycobacteri
882	18	94.7	339	1	DUS_RICCN	P23247 vibrio chol	955	18	94.7	411	1	TRUD_ARCFU	Q28596 archaeoglob
883	18	94.7	340	1	AST6_PSEAE	Q92jq6 rickettsia	956	18	94.7	411	1	CDV1_MOUSE	Q35594 mus muscullu
884	18	94.7	340	1	LINI_YEAST	P80358 pseudomonas	957	18	94.7	414	1	GAG2_DROME	P20828 drosophila
885	18	94.7	340	1	TRPD_AQUAE	P38852 saccharomyc	958	18	94.7	416	1	ODO2_ALCEU	P52993 alcaligenes
886	18	94.7	341	1	ARSA_CAEEL	O6576 aquifex aeo	959	18	94.7	417	1	GLYA_ACIRA	Q85718 acinetobact
887	18	94.7	342	1	YQ66_CAEEL	Q9zed2 rickettsia	960	18	94.7	420	1	YEO9_YEAST	P40038 saccharomyc
888	18	94.7	342	1	HRCA_THETN	P30632 caenorhabdi	961	18	94.7	425	1	ESC_DROME	Q4338 drosophila
889	18	94.7	343	1	MJY2_METJA	Q9275 caenorhabdi	962	18	94.7	425	1	ESC_DROVI	Q26458 drosophila
890	18	94.7	343	1	ABIC_LACLA	Q8rb70 thermococ	963	18	94.7	426	1	NUOF_AQUAE	O68841 aquifex aeo
891	18	94.7	344	1	YF46_METUA	Q81457 lactococcus	964	18	94.7	429	1	GLYA_METJA	Q58992 methanococ
892	18	94.7	345	1	IDHA_MACFA	Q58941 methanococ	965	18	94.7	430	1	SPAL_SHIFL	P35531 shigella fl
893	18	94.7	347	1	ARSI_HUMAN	Q28480 macaca fasc	966	18	94.7	433	1	SSAN_SALTU	P74857 salmonella
894	18	94.7	348	1	ARSI_MOUSE	Q43681 homo sapien	967	18	94.7	434	1	INGK_ECOLI	P22937 escherichia
895	18	94.7	348	1	ARSI_MOUSE	O54984 mus muscullu	968	18	94.7	434	1	TRME_UREFA	Q9prc7 ureaplasma
896	18	94.7	348	1	OTOM_RHIME	Q9zsk1 arabidopsis	969	18	94.7	438	1	PEL_LILLO	Q9prc7 ureaplasma
897	18	94.7	348	1	ODPA_RHIME	Q9r9n5 rhizobium m	970	18	94.7	440	1	YBWA_CANAL	P34948 candida alb
898	18	94.7	348	1	RECA_SINP2	P14582 synechococ	971	18	94.7	444	1	YBWA_BACSU	Q7478 saccharomyc
899	18	94.7	352	1	ARGC_SULSO	Q980x1 sulfolobus	972	18	94.7	446	1	SUB2_YEAST	P38295 saccharomyc
900	18	94.7	352	1	YPBB_BACSU	P50728 bacillus su	973	18	94.7	451	1	YB27_YEAST	P38295 saccharomyc
901	18	94.7	353	1	RECA_CHLNP	Q927e4 chlamydia p	974	18	94.7	455	1	RADA_TREFA	P38295 saccharomyc
902	18	94.7	353	1	RECA_SINP3	P74737 synechocyst	975	18	94.7	456	1	PUR8_BUCAI	P38295 saccharomyc
903	18	94.7	354	1	RECA_SINEL	Q8dh70 synechococ	976	18	94.7	460	1	UDPE_GVLO	P57351 buchnera ap
904	18	94.7	357	1	MPFE_PROMI	P53522 proteus mir	977	18	94.7	462	1	YB47_METJA	Q98166 lactococ
905	18	94.7	357	1	RECA_ANASP	P59552 anabaena sp	978	18	94.7	465	1	YHJA_ECOLI	Q58547 methanococ
906	18	94.7	358	1	RECA_ANAVA	P14167 anabaena va	979	18	94.7	467	1	ARLY_RANCA	P37197 escherichia
907	18	94.7	358	1	RNFD_HABIN	Q57288 haemophilus	980	18	94.7	469	1	SELB_METJA	P51464 rana catesb
908	18	94.7	359	1	RFPG_SALTU	P26397 salmonella	981	18	94.7	473	1	ATPB_RICGN	Q32g88 rickettsia
909	18	94.7	360	1	BUK_ENTFA	Q9rps7 enterococcu	982	18	94.7	473	1	NIFD_FRAAL	Q02452 frankia aln

Q8y3d8 Pyrobaculum  
P42590 escherichia  
P19487 xanthomonas  
P21895 bacillus su  
O52351 mycoplasma  
P57693 thermoplasma  
O56q12 caenorhabdi  
P26288 arabidopsis  
Q8y3j2 anabaena sp  
O46051 rhodospila  
P72245 rhodobacter  
P40051 saccharomyc  
O92dl2 rickettsia  
Q8abv4 c bifunction  
P26334 trypanosoma  
P38140 saccharomyc  
P49109 cavia porce  
Q8xbv3 escherichia

## ALIGNMENTS

RESULT 1  
NS2\_MYCTU  
ID NS2\_MYCTU STANDARD; PRT; 19 AA.  
AC P81136;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 30 kDa non-secretory protein 2 (fragment).  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1] NCBI\_TaxID=1773;  
RP SEQUENCE.  
RC STRAIN=H37RV;  
RA Prasad H.K., Annapurna P.S.;  
RL Submitted (DEC-1997) to Swiss-Prot.  
CC -!- SIMILARITY: STRONG, TO THE N-TERMINALS OF C.ELEGANS F20A1.4  
CC AND H.INFLUENZAE HI0967.  
CC -!- CAUTION: We are unable to find this protein in the translation of  
CC the genome of strain H37RV.  
FT NON\_TER 19 19  
SQ SEQUENCE 19 AA; 2211 MW; A8C1854BF1FF1F70 CRC64;  
Query Match 100.0%; Score 19; DB 1; Length 19;  
Best Local Similarity 100.0%; Pred.No.35;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 11 VAEF 14  
RESULT 2  
SECE\_TREPA  
ID SECE\_TREPA STANDARD; PRT; 59 AA.  
AC O83263;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Preprocein translocase secE subunit.  
GN SECE OR TP0235.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.  
CX NCBI\_TaxID=160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nichols;  
RX MEDLINE=98332770; PubMed=9665876;  
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,

RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
RA Venter J.C.;  
RT "Complete genome sequence of Treponema pallidum, the syphilis  
spirochete.";  
RL Science 281:375-388 (1998).  
CC -!- FUNCTION: Essential for protein export.  
CC -!- SUBCELLULAR LOCATION: Tail-anchored membrane protein (potential).  
CC -!- SIMILARITY: Belongs to the secE/SEC61-gamma family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE001205; AAC65223.1; -;  
DR PIR; E71349; E71349.  
DR TIGR; TP0235; -;  
DR InterPro; IPR001901; SecE.  
DR InterPro; IPR005807; SecE\_bac.  
DR Pfam; PF00584; SecE; 1.  
DR TIGRFAMs; TIGR00964; 3a0501s06; 1.  
DR PROSITE; PS01067; SECE\_SEC61G; 1.  
KW Protein transport; Translocation; Transmembrane; Complete proteome.  
FT TRANSMEM 39 59  
FT POTENTIAL.  
SQ SEQUENCE 59 AA; 6789 MW; 9AC35BA8F48B2A7F CRC64;  
Query Match 100.0%; Score 19; DB 1; Length 59;  
Best Local Similarity 100.0%; Pred.No.11e-02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
DB 12 VAEF 15  
RESULT 3  
YA90\_ARCFU  
ID YA90\_ARCFU STANDARD; PRT; 59 AA.  
AC O29175;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Hypothetical UPF0165 protein AF1090.  
GN AF1090.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370 (1997).

```

Query Match          100.0%; Score 19; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. NO. 1.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VAEF 4
      ||||
DB      27 VAEF 30

RESULT 5
RS15_HAEIN
ID      RS15_HAEIN      STANDARD;      PRT;      88 AA.
AC      P44389;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      30S ribosomal protein S15.
GN      (RPSO-A OR RPS15-A OR H11326) AND (RPSO-B OR RPS15-B OR H11468).
OS      Haemophilus influenzae.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC      Pasteurellaceae; Haemophilus.
OC      NCBI_TaxID=727;
[1]
SEQUENCE FROM N.A.
STRAIN=Rd / KW20 / ATCC 51907;
MEDLINE=95350630; PubMed=7542800;
RA      Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA      Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA      McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA      Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA      Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA      Uiterback T.R., Hanna M.C., Sprigs D.T., Saudek D.M., Brandon R.C.,
RA      Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA      Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA      Venter J.C.;
RA      "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT      Rd.";
RL      Science 269:496-512(1995).
CC      -!- FUNCTION: This protein is one of the 16S ribosomal RNA binding
CC      proteins (by similarity).
CC      -!- SIMILARITY: Belongs to the S15P family of ribosomal proteins.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      EMBL; U32812; AAC22973.1; -
CC      DR      EMBL; U32825; AAC23117.1; -
CC      DR      PIR; H64116; H64116.
CC      DR      HSSP; P05766; 1A32.
CC      DR      TIGR; H11328; -
CC      DR      TIGR; H11468; -
CC      DR      InterPro; IPR000589; Ribosomal_S15.
CC      DR      InterPro; IPR005290; Ribosomal_S15_b.
CC      DR      Pfam; PF00312; Ribosomal_S15; 1.
CC      DR      ProDom; PD157043; RS15 Bact; 1.
CC      DR      TIGRFAMs; TIGR00952; S15 bact; 1.
CC      DR      PROSITE; PS00362; RIBOSOMAL_S15; 1.
CC      KW      Ribosomal protein; rRNA-binding; Complete proteome.
FT      INIT_MET 0
FT      BY SIMILARITY.
SQ      SEQUENCE 88 AA; 10064 MW; 563BAD2B8B8A7043 CRC64;

Query Match          100.0%; Score 19; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. NO. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VAEF 4
      ||||
DB      11 VAEF 14

```

```
RESULT 6
ALB2 METKA
ID ALB2 METKA STANDARD; PRT; 93 AA.
AC Q8TW6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein Alba 2.
GN ALB2 OR MK1089.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyrales; Methanopyraceae;
OC Methanopyrus
OX NCBI_TaxID=2320;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natile D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozlovskiy S.A.;
RA "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and morphology of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC -!- FUNCTION: Binds double-stranded DNA tightly but without sequence
CC specificity. It is distributed uniformly and abundantly on the
CC chromosome, suggesting a role in chromatin architecture. However,
CC it does not significantly compact DNA (By similarity).
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- PTM: Acetylation. Deacetylation by the SIR2-homolog deacetylase
CC increases its DNA-binding affinity, thereby repressing
CC transcription. Regulation of DNA-based activities is therefore
CC achieved at the chromatin level (By similarity).
CC -!- SIMILARITY: Belongs to the archaeal histone-like Alba family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE010398; AAM02302.1; -.
CC HAMAP; MF 01122; -.
CC InterPro; IPR002775; DUF78.
CC Pfam; PF01918; DUF78; 1.
CC ProDom; PD010497; DUF78; 1.
CC TIGRfam; TIGR00285; TIGR00285; 1.
KW DNA-binding; Acetylation; Complete proteome.
FT MOD_RES 12 12 ACETYLATION (By similarity).
SQ SEQUENCE 93 AA; 10142 MW; E11D650FDF563169 CRC64;
Query Match 100.0%; Score 19; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAEF 4
Db 47 VAEF 50
RESULT 7
YF24 METJA
ID YF24 METJA STANDARD; PRT; 108 AA.
AC Q58919;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1524.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Sulton C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sulton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Overbeek R., Kirkness E.F., Tomb J.-F., Adams M.D., Reich C.I.,
RA Kerevans R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geohagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Usterback T.R., Keiley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-F., Fraser C.M., Smith H.O., Weese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- SIMILARITY: BELONGS TO THE UPF0166 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67593; AAB99550.1; -.
CC PIR; C64490; C64490.
CC TIGR; MJ1524; -.
CC InterPro; IPR003793; DUF190.
CC Pfam; PF02641; DUF190; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 108 AA; 12398 MW; D3E286AFB9D4AF8B CRC64;
Query Match 100.0%; Score 19; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAEF 4
Db 52 VAEF 55
RESULT 8
KEDA ACTSL
ID KEDA ACTSL STANDARD; PRT; 114 AA.
AC P41249;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apokedarcidin.
OS Actinomycete sp. (strain L585-6 / ATCC 53650).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales.
OX NCBI_TaxID=38989;
RN [1]
SEQUENCE.
RC STRAIN=ATCC 53650 / L585-6;
RX MEDLINE=93015257; PubMed=1399845;
RA Hofstead S.J., Matson J.A., Malacko A.R., Marquardt H.;
RT "Kedarcidin, a new chromoprotein antitumor antibiotic. II. Isolation,
RT purification and physico-chemical properties."
RL J. Antibiot. 45:1250-1254(1992).
RN [2]
CHARACTERIZATION.
RX MEDLINE=93376732; PubMed=8367457;
RA Zein N., Casazza A.M., Doyle T.W., Leet J.E., Scheoder D.R.,
RA Solomon W., Nadler S.G.;
RT "Selective proteolytic activity of the antitumor agent kedarcidin."
RL Proc. Natl. Acad. Sci. U.S.A. 90:8009-8012(1993).
RN [3]
```

```

P C STRAIN=ATCC 53650 / L585-6;
X C MDLINE=95001848; PubMed=7918359;
A Constantine K.L., Colson K.L., Wittekind M., Friedrichs M.S., Lam K.S.,
A Zain N., Tuttle J., Langley D.R., Leet J.E., Schroeder D.R., Lam K.S.,
A Farmer B.T. II, Metzler W.J., Bruccoleri R.E., Mueller L.,
T "Sequential 1H, 13C, and 15N NMR assignments and solution
L conformation of apokedarcidin."
L Biochemistry 33:11438-11452(1994).
C -I- FUNCTION: BINDS NON-COVALENTLY TO AN ENEDIYNE CHROMOPHORE WHICH IS
C THE CYTOTOXIC AND MUTAGENIC COMPONENT OF THE ANTIBIOTIC. THE
C CHROMOPHORE CLEAVES DUPLEX DNA SITE-SPECIFICALLY IN A SINGLE-
C STRANDED MANNER. THE APOPROTEIN CLEAVES PROTEINS SELECTIVELY, IN
C PARTICULAR HIGHLY BASIC HISTONES, WITH H1 PROTEINS BRING CLEAVED
C THE MORE READILY.
C -I- DOMAIN: THIS PROTEIN CONSISTS OF AN IMMUNOGLOBULIN-LIKE SEVEN-
C COMPOSED ANTIPARALLEL BETA-BARREL DOMAIN LINKED TO A SUBDOMAIN
C STRANDED OF TWO BETA-HAIRPIN RIBBONS.
C -I- SIMILARITY: Belongs to the neocarzinostatin family.
R PDB; IAKP; 3I-AUG-94.
R InterPro; IPR002186; Neocarzinostat.
R Pfam; PF00960; Neocarzinostat; 1.
R ProDom; PD012709; Neocarzinostat; 1.
W Antibiotic; DNA-binding; 3D-structure.
T DISULFID 37 47
T T STRAND 88 95
T T STRAND 4 7
T T TURN 11 13
T T STRAND 15 16
T T STRAND 18 18
T T STRAND 21 24
T T STRAND 32 33
T T STRAND 36 39
T T STRAND 46 47
T T TURN 50 51
T T STRAND 55 55
T T STRAND 61 61
T T STRAND 64 64
T T STRAND 67 67
T T STRAND 71 74
T T TURN 77 79
T T STRAND 84 87
T T STRAND 94 97
T T STRAND 100 100
T T STRAND 109 110
T T STRAND 112 114
Q SEQUENCE 114 AA; 10969 MW; 1901E2B14E4197B4 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4
b 49 VAEF 52

RESULT 9
NS_VERMO STANDARD; PRT; 115 AA.
C Q9W7R2;
T 16-OCT-2001 (Rel. 40, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
E Insulin precursor.
N INS.
S Verasper moseri (Barfin flounder).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
C Acanthomorpha; Acanthopterygii; Percormorpha; Pleuronectiformes;
C Pleuronectoides; Pleuronectidae; Verasper.
X NCBI_TaxID=98923;
N [1]

SEQUENCE FROM N.A.
Andoh T., Nagasawa H.;
"Two molecular forms of insulin from barfin flounder, Verasper moseri,
are derived from a single gene.";
Zool. Sci. 15:931-937(1998).
CC -I- FUNCTION: Insulin decreases blood glucose concentration. It
CC increases cell permeability to monosaccharides, amino acids and
CC fatty acids. It accelerates glycolysis, the pentose phosphate
CC cycle, and glycogen synthesis in liver.
CC -I- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: Belongs to the insulin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB029318; BAA82315.1; -.
CC HSP; P01315; IMPJ.
CC InterPro; IPR004825; Ins/IGF/relax.
CC Pfam; PF00049; Insulin; 1.
CC PRINTS; PR00277; INSULINB.
CC SMART; SM00078; IIGF; 1.
CC PROSITE; PS00262; INSULIN; 1.
CC Insulin family; Hormone; Glucose metabolism; Signal.
CC SIGNAL 1 22 BY SIMILARITY.
CC CHAIN 23 53 INSULIN B CHAIN.
CC PROPEP 56 92 C PEPTIDE.
CC CHAIN 95 115 INSULIN A CHAIN.
CC DISULFID 32 101 INTERCHAIN (BY SIMILARITY).
CC DISULFID 44 114 INTERCHAIN (BY SIMILARITY).
CC DISULFID 100 105 BY SIMILARITY.
SQ SEQUENCE 115 AA; 12608 MW; 7EA2A5B568DEDBB CRC64;

Query Match 100.0%; Score 19; DB 1; Length 115;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
Db 79 VAEF 82

RESULT 10
INS_LOPPI STANDARD; PRT; 116 AA.
AC P01341;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Insulin precursor.
GN INS.
OS Lophius piscatorius (Allmouth goosefish) (Anglerfish), and
OS Lophius americanus (American goosefish) (Anglerfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Lophiiformes; Lophiidae; Lophius.
OX NCBI_TaxID=8074, 8073;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=L.americanus;
RX MEDLINE=81056434; PubMed=7001633;
RA Hobart P.M., Shen L.-P., Crawford R., Pictet R.L., Rutter W.J.;
RT "Comparison of the nucleic acid sequence of anglerfish and mammalian
RT insulin mRNA's from cloned cdna's.";
RL Science 210:1360-1363(1980).
RN [2]
RP SEQUENCE OF 25-54 AND 96-116.

```

RC SPECIES=L.piscatorius;  
 RX MEDLINE=70036620; PubMed=5389298;  
 RA Neumann P.A., Koldenhof M., Humbel R.E.;  
 RT "Amino acid sequence of insulin from the angler fish (Lophius  
 piscatorius)".  
 RL Hoppe-Sevler's Z. Physiol. Chem. 350:1286-1288(1969).  
 CC -|- FUNCTION: Insulin decreases blood glucose concentration. It  
 increases cell permeability to monosaccharides, amino acids and  
 fatty acids. It accelerates glycolysis, the pentose phosphate  
 cycle, and glycogen synthesis in liver.  
 CC -|- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
 disulfide bonds.  
 CC -|- SUBCELLULAR LOCATION: Secreted.  
 CC -|- SIMILARITY: Belongs to the insulin family.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC EMBL; V00634; CAA23907.1; --  
 DR PIR; A01608; IPAF.  
 DR HSP; P01308; ILPH.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULIN.  
 DR SMART; SM00078; ILGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 53 INSULIN B CHAIN.  
 FT PROPEP 56 93 C PEPTIDE.  
 FT CHAIN 96 116 INSULIN A CHAIN.  
 FT DISULFID 32 102 INTERCHAIN.  
 FT DISULFID 44 115 INTERCHAIN.  
 FT DISULFID 101 106  
 SQ SEQUENCE 116 AA; 12737 MW; C686F8EF8183BEFE CRC64;  
 Query Match 100.0%; Score 19; DB 1; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 DB 80 VAEF 83  
 RESULT 11  
 U279 DROME STANDARD; PRT; 119 AA.  
 AC Q9V8F3;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE UF0279 protein CG14505.  
 GN CG14505.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 ABRI J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burits K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Fabios B., Delcher A., Deng C., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostel C., Gong F., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster".  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RT "A Drosophila full-length cDNA resource".  
 CC Genome Biol. 3:RESEARCH080.1-RESEARCH080.8(2002).  
 CC -|- SIMILARITY: Belongs to the UF0279 family.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC EMBL; AE003800; AAF57714.1; --  
 DR EMBL; AY084100; AAL69838.1; --  
 DR FlyBase; FBGN0034327; CG14505.  
 DR InterPro; IPR007967; DUF727.  
 DR Pfam; PF05303; DUF727; 1.  
 SQ SEQUENCE 119 AA; 13534 MW; 83FA23FCCC389AA CRC64;  
 Query Match 100.0%; Score 19; DB 1; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 DB 33 VAEF 36  
 RESULT 12  
 GTR2\_PIG STANDARD; PRT; 120 AA.  
 ID GTR2\_PIG  
 AC 062786;  
 DT 28-FEB-2003 (Rel. 41, Created)

T 28-FEB-2003 (Rel. 41, Last sequence update)  
 T 28-FEB-2003 (Rel. 41, Last annotation update)  
 E Solute carrier family 2, facilitated glucose transporter, member 2  
 E (Glucose transporter type 2, liver) (fragment).  
 N SLC2A2 OR GLUT2.  
 S Sus scrofa (Pig).  
 C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 C Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 X NCBI\_TaxID=9823;  
 N [1]  
 P SEQUENCE FROM N.A.  
 A Catty J.M., Young R.F., Fallavollita J.A.;  
 L Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 C -1- FUNCTION: Facilitative glucose transporter. This isoform likely  
 C mediates the bidirectional transfer of glucose across the plasma  
 C membrane of hepatocytes and is responsible for uptake of glucose  
 C by beta cells; may comprise part of the glucose-sensing  
 C mechanism of the beta cell. May also participate with the  
 C Na(+)/glucose cotransporter in the transcellular transport of  
 C glucose in the small intestine and kidney (By similarity).  
 C -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 C -1- SIMILARITY: Belongs to the sugar transporter family. Glucose  
 C transporter subfamily.  
 C  
 C This SWISS-PROT entry is copyright. It is produced through a collaboration  
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 C the European Bioinformatics Institute. There are no restrictions on its  
 C use by non-profit institutions as long as its content is in no way  
 C modified and this statement is not removed. Usage by and for commercial  
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 C  
 C EMBL; AF054835; AAC12737.1; -.  
 R InterPro; IPR007114; MFS.  
 R InterPro; IPR005828; Sub\_transporter.  
 R InterPro; IPR005829; Sug\_transporter.  
 R InterPro; IPR003663; Sugar\_transp.  
 R Pfam; PF00083; sugar tr; 1.  
 R PRINTS; PR00171; SUGTRANSFPORT.  
 R PROSITE; PS00850; MFS; 1.  
 R PROSITE; PS00216; SUGAR\_TRANSPORT\_1; PARTIAL.  
 R PROSITE; PS00217; SUGAR\_TRANSPORT\_2; PARTIAL.  
 W Transmembrane; Sugar transport; Transport; Multigene family.  
 N NON TER 1 1  
 T TRANSMEM 2 22 9 (POTENTIAL).  
 T DOMAIN 23 26 EXTRACELLULAR (POTENTIAL).  
 T TRANSMEM 27 47 10 (POTENTIAL).  
 T DOMAIN 48 56 CYTOPLASMIC (POTENTIAL).  
 T TRANSMEM 57 77 11 (POTENTIAL).  
 T DOMAIN 78 84 EXTRACELLULAR (POTENTIAL).  
 T TRANSMEM 85 105 12 (POTENTIAL).  
 T DOMAIN 106 120 CYTOPLASMIC (POTENTIAL).  
 T NON TER 120 120  
 Q SEQUENCE 120 AA; 13503 MW; D5F73168BDF03203 CRC64;  
 Query Match 100.0%; Score 19; DB 1; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 1 VAEF 4  
 b 46 VAEF 49  
 RESULT 13  
 D\_R18E PYRAE STANDARD; PRT; 122 AA.  
 C Q8ZYQ2; 2003 (Rel. 41, Created)  
 T 28-FEB-2003 (Rel. 41, Last sequence update)  
 T 28-FEB-2003 (Rel. 41, Last annotation update)  
 E 50S ribosomal protein L18e.  
 N RPL18E OR PAE0572.

OS Pyrobaculum aerophilum.  
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
 OC Thermoproteaceae; Pyrobaculum.  
 OX NCBI\_TaxID=13773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
 RX MEDLINE=21664397; PubMed=11792869;  
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller J.H.;  
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
 RT aerophilum".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
 CC -1- SIMILARITY: Belongs to the L18E family of ribosomal proteins.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE009779; AAL62941.1; -.  
 DR HAMAP; MF 00329; -; 1.  
 DR InterPro; IPR001196; Ribosomal\_L15.  
 DR InterPro; IPR000039; Ribosomal\_L18e.  
 DR Pfam; PF00356; L15; 1.  
 DR PROSITE; PS01106; RIBOSOMAL\_L18E; FALSE\_NEG.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 122 AA; 13252 MW; 88F3DB732C4E0394 CRC64;  
 Query Match 100.0%; Score 19; DB 1; Length 122;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 Db 35 VAEF 38  
 RESULT 14  
 Y670\_PASMU STANDARD; PRT; 124 AA.  
 ID Y670\_PASMU  
 AC Q9CMY0;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein PM0670 precursor.  
 GN PM0670.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Pasteurella.  
 OX NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Em70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida Em70.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 CC -1- SIMILARITY: Belongs to the cytochrome b562 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE006103; AAK02754.1; -.  
 DR PIRSF; PIRSF000029; Cytochrome\_b562; 1.

KW Hypothetical protein; Signal; Complete proteome.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 124 HYPOTHETICAL PROTEIN PM0670.  
 SQ SEQUENCE 124 AA; 13746 MW; D7B2B485C7B51B9A CRC64;  
 Query Match 100.0%; Score 19; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 DB 45 VAEF 48  
 RESULT 15  
 ILBP\_MOUSE  
 ID ILBP\_MOUSE STANDARD; PRT; 127 AA.  
 AC P51162;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Gastrotropin (GT) (ileal lipid-binding protein) (ILBP).  
 DE FAPB6 OR ILBP.  
 DE Mus musculus (Mouse).  
 DE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 DE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 DX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DBA/2J; TISSUE=Liver;  
 RX MEDLINE=94375529; PubMed=8089185;  
 RA Crossman M.W., Hauf S.M., Gordon J.I.;  
 RT "The mouse ileal lipid-binding protein gene: a model for studying  
 axial patterning during gut morphogenesis.";  
 RL J. Cell Biol. 126:1547-1564 (1994).  
 CC -!- FUNCTION: ILEAL PROTEIN WHICH STIMULATES GASTRIC ACID AND  
 PEPINOGEN SECRETION. SEEMS TO BE ABLE TO BIND TO BILE SALTS AND  
 BILIRUBIN.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to the fatty-acid binding protein (FABP)  
 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; U00938; ARC27352.1; -.  
 DR PIR; A54797; A54797.  
 DR HSP; F10289; IEAL.  
 DR MGD; MGI:96565; Fapb6.  
 DR InterPro; IPR000463; Fatty\_acid\_BP.  
 DR InterPro; IPR000566; Lipocln\_cytFAPB.  
 DR Pfam; PF00061; lipocalin; 1.  
 DR PRINTS; PR00178; FATTYACIDBP.  
 DR PROSITE; PS00214; FAPB; 1.  
 KW Transport; Lipid-binding; Acetylation.  
 FT INIT\_MET 0  
 FT MOD\_RES 1  
 FT BY SIMILARITY.  
 FT ACETYLATION (BY SIMILARITY).  
 SQ SEQUENCE 127 AA; 14355 MW; 0690BDD0A9CD922 CRC64;  
 Query Match 100.0%; Score 19; DB 1; Length 127;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 DB 91 VAEF 94

RESULT 16  
 SPEH\_THEME  
 ID SPEH\_THEME STANDARD; PRT; 130 AA.  
 AC Q9WZC3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50) (AdoMetDC)  
 DE (SAMDC) [Contains: S-adenosylmethionine decarboxylase beta chain; S-  
 adenosylmethionine decarboxylase alpha chain].  
 DE SPEH OR TM0655.  
 GN Thermotoga maritima.  
 OS Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
 OC NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB / DSM 3109 / ATCC 43589;  
 RX MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Steward A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 genome sequence of Thermotoga maritima.";  
 RL Nature 399:323-329 (1999).  
 CC -!- FUNCTION: Decarboxylation of S-adenosylmethionine provides the  
 aminopropyl moiety required for spermidine biosynthesis from  
 putrescine (By similarity).  
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine = (S)-deoxy-5-  
 adenosyl(3-aminopropyl) methylsulfonium salt + CO(2).  
 CC -!- COFACTOR: Pyruvoyl group (By similarity).  
 CC -!- SIMILARITY: Belongs to the prokaryotic AdoMetDC family. Subfamily  
 1.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; AE001739; AAD35739.1; -.  
 DR PIR; D72348; D72348.  
 DR TIGR; TM0655; -.  
 DR HAMAP; MF\_00464; -; 1.  
 DR InterPro; IPR003826; SAMDC.  
 DR Pfam; PF02675; AdoMetDC; 1.  
 DR Spermidine biosynthesis; Lyase; Decarboxylase; Zymogen; Pyruvate;  
 KW Complete proteome.  
 FT CHAIN 1 62 S-ADENOSYLMETHIONINE DECARBOXYLASE BETA  
 FT CHAIN 63 130 S-ADENOSYLMETHIONINE DECARBOXYLASE ALPHA  
 FT SITE 62 63 CHAIN (BY SIMILARITY).  
 FT MOD\_RES 63 63 CLEAVAGE (NONHYDROLYTIC) (BY SIMILARITY).  
 FT CONVERTED TO A PYRUVOYL GROUP (BY  
 SIMILARITY).  
 SQ SEQUENCE 130 AA; 14785 MW; 7659FE20A2019928 CRC64;  
 Query Match 100.0%; Score 19; DB 1; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 DB 9 VAEF 12  
 RESULT 17  
 Y194\_PYRAB  
 ID Y194\_PYRAB STANDARD; PRT; 131 AA.

Q9V280;  
 16-OCT-2001 (Rel. 40, Created)  
 16-OCT-2001 (Rel. 40, Last sequence update)  
 10-OCT-2003 (Rel. 42, Last annotation update)  
 Hypothetical UPF0146 protein PYAB01940.  
 PYAB01940 OR PAB2224.  
 Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;  
 Pyrococcus abyssi.  
 NCBI\_TaxID=29292;  
 [1]\_SEQUENCE FROM N.A.  
 STRAIN=GES / Orsay;  
 MEDLINE=22511545; PubMed=12622808;  
 Cohen G.N., Barbe V., Flamant D., Galperin M., Heilig R., Lecompte O.,  
 Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,  
 Van der Oost J., Weissbach J., Zivanovic Y., Forterre P.;  
 "An integrated analysis of the genome of the hyperthermophilic  
 archaeon Pyrococcus abyssi.";  
 Mol. Microbiol. 47:1495-1512(2003).  
 -!- SIMILARITY: Belongs to the UPF0146 family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; AJ248283; CAB49118.1; -;  
 F1R; G75208; G75208.  
 HAMAP; MF\_00341; -; 1.  
 InterPro; IPR005353; UPF0146.  
 Pfam; PF03686; UPF0146; 1.  
 PIRSF; PIRSF016725; UCP016725; 1.  
 ProDom; PD021130; UPF0146; 1.  
 Hypothetical protein; Complete proteome.  
 SEQUENCE 131 AA; 14629 MW; D13F378187F832C3 CRC64;  
 -----

Query Match 100.0%; Score 19; DB 1; Length 131;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 VAEF 4  
 4 VAEF 7

RESULT 18  
 ISB HALNI  
 KISE HALNI STANDARD; PRT; 133 AA.  
 Q9HRM5;  
 16-OCT-2001 (Rel. 40, Created)  
 16-OCT-2001 (Rel. 40, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)  
 (lumazine synthase) (Riboflavin synthase beta chain).  
 R1EH OR R1BE OR VNG0630G.  
 Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).  
 Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 Halobacteriaceae; Halobacterium.  
 NCBI\_TaxID=64091;  
 [1]\_SEQUENCE FROM N.A.  
 MEDLINE=20504483; PubMed=11016950;  
 Ng M.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
 Shukla H.D., Lesky S.R., Baliga N.S., Thorsen V., Sirogna J.,  
 Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 Maddocks D.G., Jablonski P.E., Krebs M.P., Angelvine C.M., Dale H.,  
 Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
 Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,

RA Eberhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC -!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex  
 catalyzing the formation of riboflavin from 5-amino-6-(1',D)-  
 ribityl-amino-2,4(1H,3H)-pyrimidinedione and L-3,4-dihydroxy-2-  
 butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit  
 catalyzes the condensation of 5-amino-6-(1',D)-ribityl-amino-  
 2,4(1H,3H)-pyrimidinedione with L-3,4-dihydroxy-2-butanone-4-  
 phosphate yielding 6,7-dimethyl-8-lumazine (By similarity).  
 CC CATABOLIC ACTIVITY: 6,7-dimethyl-8-(1-D-ribityl)lumazine =  
 CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.  
 CC -!- SIMILARITY: Belongs to the DMRL synthase family.  
 CC -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; AE005011; AAG19133.1; -;  
 F1R; A84221; A84221.  
 HSP; O65529; 1HQK.  
 HAMAP; MF\_00178; -; 1.  
 InterPro; IPR002180; DMRL synthase.  
 Pfam; PF00885; DMRL synthase; 1.  
 ProDom; PD003664; DMRL synthase; 1.  
 KW Riboflavin biosynthesis; Transferase; Complete proteome.  
 SQ SEQUENCE 133 AA; 13719 MW; B9724DC24650D6CD CRC64;  
 -----

Query Match 100.0%; Score 19; DB 1; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 VAEF 4  
 8 VAEF 11

RESULT 19  
 YBGC ECOLI  
 ID YBGC ECOLI STANDARD; PRT; 134 AA.  
 AC P08399;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protein ybgC.  
 GN YBGC OR B0736 OR C0815 OR Z0904 OR ECS0771 OR SF0561 OR S0574.  
 OS Escherichia coli.  
 OS Escherichia coli O6.  
 OS Escherichia coli O157:H7, and  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OC NCBI\_TaxID=562, 217992, 83334, 623;  
 RN [1]\_SEQUENCE FROM N.A.  
 RP SPECIES=E.coli;  
 RC MEDLINE=8722192; PubMed=3294803;  
 RA Sun T.-P., Webster R.E.;  
 RT "Nucleotide sequence of a gene cluster involved in entry of E. coli  
 and single-stranded DNA of infecting filamentous bacteriophages into  
 Escherichia coli.";  
 RT J. Bacteriol. 169:2667-2674(1987).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP SPECIES=E.coli; STRAIN=K12 / MG1655;  
 RC MEDLINE=97426617; PubMed=9278503;  
 RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.,  
 RA "The complete genome sequence of *Escherichia coli* K-12.";  
 RL Science 277:1453-1474 (1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sampei G., Seki Y., Tagami H., Takenoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiuchi T.,  
 RA "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome  
 RT corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155 (1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Raeko D., Buckles E.L., Liou S.-P., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Ferna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RA "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic *Escherichia coli*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postfai G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,  
 RA Gotzbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RA "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";  
 RL Nature 409:529-533 (2001).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RA "Complete genome sequence of enterohemorrhagic *Escherichia coli*  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:111-22 (2001).  
 RN [7]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RC SPECIES=E.coli; STRAIN=K12 / MG1655;  
 RA Kim K., Allen E., Araujo R., Aparicio A.M., Botstein D.,  
 RA Cherry M., Chung E., Dietrich F., Duncan M., Federspiel N.,  
 RA Kalman S., Komp C., Lashkari D., Lew H., Lin D., Namath A.,  
 RA Oefner P., Davis R.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
 RX MEDLINE=2272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao B., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.;  
 RA "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity  
 RT through comparison with genomes of *Escherichia coli* K12 and O157.";  
 RL Nucleic Acids Res. 30:4432-4441 (2002).  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 RX MEDLINE=22590274; PubMed=12704152;  
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky D.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.;  
 RA "Complete genome sequence and comparative genomics of *Shigella*  
 RT flexneri serotype 2a strain 2457T.";  
 RL Infect. Immun. 71:2775-2786 (2003).  
 RN [10]  
 RP IDENTIFICATION BY MASS SPECTROMETRY.  
 RC SPECIES=E.coli;  
 RX MEDLINE=99420866; PubMed=10493123;  
 RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;  
 RA "Enrichment of low abundance proteins of *Escherichia coli* by  
 RT hydroxyapatite chromatography.";  
 RL Electrophoresis 20:2181-2195 (1999).  
 CC -1- SIMILARITY: BELONGS TO THE 4-HYDROXYBENZYL-COA THIOESTERASE  
 CC FAMILY. STRONG, TO H-INFLUENZAE H10386.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M16489; AAA3918.1; -;  
 CC EMBL; AE000177; AAC73830.1; -;  
 CC EMBL; D90713; BAA35402.1; -;  
 CC EMBL; AE016757; AAN79488.1; -;  
 CC EMBL; AE005252; AAG55072.1; -;  
 CC EMBL; AP002553; BAB34194.1; -;  
 CC EMBL; U30934; AAA74398.1; -;  
 CC EMBL; AE015086; AAM42205.1; -;  
 CC EMBL; AE016979; AAP16078.1; -;  
 CC EMBL; A25980; WMEC15.  
 CC PIR; C90725; C90725.  
 CC PIR; D85576; D85576.  
 CC EcoGene; EG1110; ybgC.  
 CC InterPro; IPR008272; 4HBCOA\_thiostrase.  
 CC InterPro; IPR006684; 4HBCOA\_thiostrase.  
 CC InterPro; IPR006683; Thioestr\_sufp.  
 CC Pfam; PF03063; 4HBT; 1.  
 CC TIGRFAms; TIGR00051; TIGR00051; 1.  
 CC PROSITE; PS01329; 4HBCOA\_THIOESTERASE; 1.  
 CC Hydrolase; Complete proteome.  
 CC ACT SITE 18  
 CC ACT SITE 18 BY SIMILARITY.  
 CC SEQUENCE 134 AA; 15562 MW; C44582B6EC3BE989 CRC64;  
 CC -----  
 CC Query Match 100.0%; Score 19; DB 1; Length 134;  
 CC Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 CC Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC QV 1 VAEF 4  
 CC 129 VAEF 132  
 CC -----  
 CC Db 129 VAEF 132  
 CC -----  
 CC RESULT 20  
 CC RISB METH STANDARD; PRT; 139 AA.  
 CC ID RISB METH  
 CC AC 027443;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE 6,7-dimethyl-8-ribitylmuramic acid synthase (EC 2.5.1.9) (DMRL synthase)  
 CC (Lumazine synthase) (Riboflavin synthase beta chain).  
 CC GN RIBH OR MTH1390.  
 CC OS Methanobacterium thermoautotrophicum.  
 CC OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
 CC OC Methanobacteriaceae; Methanothermobacter.

```

X NCBI_TaxID=187420;
X [1]
X SEQUENCE FROM N.A.
X STRAIN=VC-16;
X MEDLINE=98037514; PubMed=9371463;
X Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
X Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
X Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
X Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
X Jiwni N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
X McDougall S., Shmer G., Goyal A., Pietrowski S., Church G.M.,
X Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
X "Complete genome sequence of Methanobacterium thermoautotrophicum
X deltaH: functional analysis and comparative genomics.";
X J. Bacteriol. 179:7135-7155 (1997).
X -!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
X catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
X ribityl-amino-2,4 (1H,3H)-pyrimidinone and L-3,4-dihydroxy-2-
X butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
X catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-
X 2,4 (1H,3H)-pyrimidinone with L-3,4-dihydroxy-2-butanone-4-
X phosphate yielding 6,7-dimethyl-8-lumazine (By similarity).
X -!- CATALYTIC ACTIVITY: 2,6,7-dimethyl-8-(1'-D-ribityl)lumazine =
X riboflavin + 4-(1'-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
X -!- PATHWAY: Riboflavin biosynthesis; last step.
X -!- SIMILARITY: Belongs to the DMRL synthase family.
X
X This SWISS-PROT entry is copyright. It is produced through a collaboration
X between the Swiss Institute of Bioinformatics and the EMBL outstation -
X the European Bioinformatics Institute. There are no restrictions on its
X use by non-profit institutions as long as its content is in no way
X modified and this statement is not removed. Usage by and for commercial
X entities requires a license agreement (See http://www.isb-sib.ch/announce/
X or send an email to license@isb-sib.ch).
X
X-----
X EMBL; AE001064; AAB90658.1; -
X PIR; B69323; B69323.
X TIGR; AF0586; -.
X KW Hypothetical protein; Complete proteome.
X SQ SEQUENCE 140 AA; 16175 MW; D3B1FE3A62A05901 CRC64;
X
X Query Match 100.0%; Score 19; DB 1; Length 140;
X Best Local Similarity 100.0%; Pred. NO. 2.5e+02; Indels 0; Gaps 0;
X Matches 4; Conservative 0; Mismatches 0;
X
X QY 1 VAEF 4
X ||||
X Db 65 VAEF 68
X
X RESULT 22
X RIBS ARCFU STANDARD; PRT; 143 AA.
X ID RIBS ARCFU STANDARD; PRT; 143 AA.
X AC Q28152;
X DT 16-OCT-2001 (Rel. 40, Created)
X DT 16-OCT-2001 (Rel. 40, Last sequence update)
X DT 28-FEB-2003 (Rel. 41, Last annotation update)
X DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)
X (lumazine synthase) (Riboflavin synthase beta chain).
X GN RIBH OR RIBE OR AF2128.
X OS Archaeoglobus fulgidus.
X OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
X OC Archaeoglobaceae; Archaeoglobus.
X OX NCBI_TaxID=2234;
X RN [1]
X RP SEQUENCE FROM N.A.
X RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
X RX MEDLINE=98049343; PubMed=9389475;
X RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
X RA Richardson D.L., Kerlavage A.R., Graham D.E., Hickey E.K., Kyrpides N.C.,
X RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
X RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
X RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
X RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
X RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
X RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
X RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
X RA Venter J.C.;
X RT "The complete genome sequence of the hyperthermophilic, sulphate-
X reducing archaeon Archaeoglobus fulgidus.";
X RL Nature 390:364-370 (1997).
X CC -!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
X catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
X ribityl-amino-2,4 (1H,3H)-pyrimidinone and L-3,4-dihydroxy-2-
X butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
X
X Y 1 VAEF 4
X ||||
X b 10 VAEF 13
X
X RESULT 21
X S86 ARCFU STANDARD; PRT; 140 AA.
X C Q29659;
X T 16-OCT-2001 (Rel. 40, Created)
X T 16-OCT-2001 (Rel. 40, Last sequence update)
X T 16-OCT-2001 (Rel. 40, Last annotation update)
X E Hypothetical protein AF0586.
X S AF0586.
X N Archaeoglobus fulgidus.
X C Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
X C Archaeoglobaceae; Archaeoglobus.
X X NCBI_TaxID=2234;
X [1]
X P SEQUENCE FROM N.A.
X C STRAIN=VC-16 / DSM 4304 / ATCC 49558;
X X MEDLINE=98049343; PubMed=9389475;
X A Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

```

CC catalyzes the condensation of 5-amino-6-(1'-D-riboityl)-amino-  
 CC 2,4(1H,3H)-pyrimidinone with L-3,4-dihydroxy-2-butanone-4-  
 CC phosphate yielding 6,7-dimethyl-8-lumazine (By similarity).  
 CC -!- CATALYTIC ACTIVITY: 2,6,7-dimethyl-8-(1-D-riboityl)lumazine =  
 CC riboflavin + 4-(1-D-riboitylamino)-5-amino-2,6-dihydroxypyrimidine.  
 CC -!- PATHWAY: Riboflavin biosynthesis; last step.  
 CC -!- SIMILARITY: Belongs to the DMRL synthase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF000957; AAB89124.1; -.  
 CC PIR; H69515; H69515.  
 CC HSSP; P11998; 1RVV.  
 CC TIGR; AF2128; -.  
 CC HAMAP; MF 00178; -; 1.  
 CC InterPro; IPR002180; DMRL\_synthase.  
 CC Pfam; PF00885; DMRL\_synthase; 1.  
 CC ProDom; PD003664; DMRL\_synthase; 1.  
 CC TIGRFAMs; TIGR00114; ribH; 1.  
 CC Riboflavin biosynthesis; transferase; Complete proteome.  
 CC KW Riboflavin biosynthesis; transferase; Complete proteome.  
 CC SEQUENCE 143 AA; 15645 MW; BBA3141206905410 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 D5 10 VAEF 13

## RESULT 23

ID RL13 HALMA STANDARD; PRT; 145 AA.  
 AC P29198;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE 50S ribosomal protein L13p (Hmal13).  
 GN RPL13P.  
 OS Haloarcula marismortui (Haloacterium marismortui).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Haloarcula.  
 OX NCBI\_TaxID=2238;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=92105119; PubMed=1940597;  
 RX Kroeber W.J., Arndt E.;  
 RA "Haloarcula S9 operon. Three ribosomal protein genes are  
 RT cotranscribed with genes encoding a tRNA(Leu), the enolase, and a  
 RT putative membrane protein in the archaeobacterium Haloarcula  
 RT (Haloacterium) marismortui".  
 RL J. Biol. Chem. 266:24573-24579 (1991).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RC STRAIN=ATCC 43049;  
 RX MEDLINE=20296344; PubMed=10937989;  
 RA Ban N., Nissen P., Hansen J., Moore P.B., Steitz T.A.;  
 RT "The complete atomic structure of the large ribosomal subunit at 2.4  
 RT A resolution".  
 RL Science 289:905-920(2000).  
 CC -!- SIMILARITY: Belongs to the L13p family of ribosomal proteins.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL; M76567; AAA73097.1; -.  
 CC PIR; B41715; B41715.  
 CC PDB; 1PFK; 26-SEP-01.  
 CC PDB; 1K8A; 19-JUL-02.  
 CC PDB; 1K9M; 19-JUL-02.  
 CC PDB; 1KD1; 19-JUL-02.  
 CC PDB; 1MIK; 23-AUG-02.  
 CC InterPro; IPR005822; Ribosomal L13.  
 CC InterPro; IPR005755; Ribosomal L13e/a.  
 CC Pfam; PF00572; Ribosomal L13; 1.  
 CC TIGRFAMs; TIGR01077; L13\_A\_E; 1.  
 CC PROSITE; PS00783; RIBOSOMAL\_L13; 1.  
 CC KW Ribosomal protein; 3D-structure.  
 CC SEQUENCE 145 AA; 16228 MW; 069CE66662AE3BC CRC64;

Query Match 100.0%; Score 19; DB 1; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 D5 3 VAEF 6

## RESULT 24

ID CLM4 MOUSE STANDARD; PRT; 148 AA.  
 AC Q9JMB3; Q9CR31; Q9DIE9;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Calmodulin 4 (Calcium-binding protein Dd112).  
 GN CALM4 OR Dd112.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ikegawa S., Nakamura Y.;  
 RT "Dd112, a novel mouse gene implicated in the early stage of ectopic  
 RT ossification".  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Tongue;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischnann W., Gaasterland T., Gisi C., King B., Kochiwa H.,  
 RA Kuernl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gusticich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki K., Sato K., Schoenbach C., Seya I., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection".  
 RL Nature 403:685-690(2001).  
 CC -!- FUNCTION: Implicated in the early stage of ectopic ossification.

1- SIMILARITY: Contains 3 EF-hand calcium-binding domains.  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; A036744; BAA95412.1; -  
EMBL; AK009556; BAB26608.1; -  
EMBL; AK009664; BAB26425.1; -  
EMBL; AK003648; BAB22914.1; -  
HSSP; P02593; 1CDM.  
MGD; MGI:1931464; Calm4.  
InterPro; IPR002048; EF-hand.  
InterPro; IPR001125; Recoverin.  
Pfam; PF00036; ehand; 4.  
PRINTS; PR00450; RECOVERIN.  
ProDom; PD000012; EF-hand; 2.  
SMART; SM00054; EPH; 4.  
PROSITE; PS00018; EF\_HAND; 3.  
Calcium-binding; Repeat.  
CA BIND 21 32 EF-HAND 1 (POTENTIAL).  
T CA BIND 21 32 EF-HAND 2 (POTENTIAL).  
T CA BIND 57 68 EF-HAND 3 (POTENTIAL).  
T CA BIND 93 104 E -> V (IN REF. 2; BAB22914).  
T CONFLICT 9 9 E -> V (IN REF. 2; BAB22914).  
T CONFLICT 124 124 M -> V (IN REF. 2; BAB22914).  
T CONFLICT 146 148 VEN -> I (IN REF. 1).  
Q SEQUENCE 148 AA; 16767 MW; 2AE22BC738AD3F9D CRC64;  
Query Match 100.0%; Score 19; DB 1; Length 148;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 VAEF 4  
b 10 VAEF 13  
RESULT 25  
ORF PYRMO STANDARD; PRT; 148 AA.  
C P20296;  
T 01-FEB-1991 (Rel. 17, Created)  
T 01-FEB-1991 (Rel. 17, Last sequence update)  
T 16-OCT-2001 (Rel. 40, Last annotation update)  
E Hypothetical protein in GAPDH 3'region (ORF X) (Fragment).  
S Pyrococcus woesei.  
C Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
C Pyrococcus.  
X NCBI\_TaxID=2262;  
N [1]  
P SEQUENCE FROM N.A.  
C STRAIN=DSM 3773;  
X MEDLINE=80330536; PubMed=2165475;  
A Zwickl P., Fabry S., Bogedain C., Haas A., Hensel R.;  
T "Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic  
T archaeabacterium Pyrococcus woesei: characterization of the enzyme,  
T cloning and sequencing of the gene, and expression in Escherichia  
T coli".  
L J. Bacteriol. 172:4329-4338 (1990).  
R PIR; S10655; S10655.  
W Hypothetical protein.  
T NON TER 1  
Q SEQUENCE 148 AA; 15256 MW; 8C7E51A072DF5D50 CRC64;  
Query Match 100.0%; Score 19; DB 1; Length 148;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 VAEF 4

Db 115 VAEF 118  
RESULT 26  
RIBB PYRAE STANDARD; PRT; 150 AA.  
AC Q8ZTE3;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)  
DE (lumazine synthase) (Riboflavin synthase beta chain).  
GN RIBH OR PAE3296.  
OS Pyrobaculum aerophilum.  
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
OC Thermoproteaceae; Pyrobaculum.  
OX NCBI\_TaxID=13773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
RX MEDLINE=21664397; PubMed=11792869;  
RA Ritz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
RA Miller J.H.;  
RT genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
RT aerophilum".  
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989 (2002).  
CC -1- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex  
CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-  
CC ribityl-amino-2,4(1H,3H)-pyrimidinone and L-3,4-dihydroxy-2-  
CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit  
CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-  
CC amino-2,4(1H,3H)-pyrimidinone with L-3,4-dihydroxy-2-  
CC butanone-4-phosphate yielding 6,7-dimethyl-8-lumazine (By  
CC similarity).  
CC -1- CATALYTIC ACTIVITY: 2,6,7-dimethyl-8-(1'-D-ribityl)lumazine =  
CC riboflavin + 4-(1'-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.  
CC -1- PATHWAY: Riboflavin biosynthesis; last step.  
CC -1- SIMILARITY: Belongs to the DMRL synthase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AR009923; RAL64819.1; -  
CC HAMAP; MF 00178; -; 1.  
DR InterPro; IPR002180; DMRL synthase.  
DR Pfam; PF00885; DMRL synthase; 1.  
DR ProDom; PD003664; DMRL synthase; 1.  
DR TIGRFAMs; TIGR00114; rlbH; 1.  
KW Riboflavin biosynthesis; Transferase; Complete proteome.  
SQ SEQUENCE 150 AA; 16474 MW; 73751C1363A8AD8D CRC64;  
Query Match 100.0%; Score 19; DB 1; Length 150;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
Db 8 VAEF 11  
RESULT 27  
SODC PRIGL STANDARD; PRT; 152 AA.  
ID SODC PRIGL  
AC P11418;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).
DS Prionace glauca (Blue shark).
DC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
CC Elasmobranchii; Galeomorphii; Galeoides; Carcharhiniformes;
CC Carcharhinidae; Prionace.
DX NCBI_TaxID=7815;
ZN [1]
RP SEQUENCE.
RX MEDLINE=89290032; PubMed=2500367;
RA Calabrese L., Polticelli F., O'Neill P., Galtieri A., Barra D.,
ZA Schinina M.E., Bossa F.;
RT "Substitution of arginine for lysine 134 alters electrostatic
KT parameters of the active site in shark Cu,Zn superoxide dismutase.";
RL FEBS Lett. 250:49-52(1989).
YC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
XR PIR; S04623; S04623.
XR HSSP; P00442; 1CBJ.
XR InterPro; IPR001424; SOD_CU_ZN.
XR Pfam; PF00080; sdcu; 1.
XR PRINTS; PR00068; CUZNDISMTASE.
XR ProDom; PD000469; SOD_CU_ZN; 1.
XR PROSITE; PS00087; SOD_CU_ZN; 1.
XR PROSITE; PS00332; SOD_CU_ZN; 2; 1.
CW Antioxidant; Oxidoreductase; Metal-binding; Copper; Zinc.
FT METAL 44 44
FT METAL 46 46 COPPER.
FT METAL 61 61 COPPER AND ZINC.
FT METAL 69 69 ZINC.
FT METAL 78 78 ZINC.
FT METAL 81 81 ZINC.
FT METAL 118 118 COPPER.
FT METAL 118 118 BY SIMILARITY.
FT DISULFID 55 144
FT SEQUENCE 152 AA; 15840 MW; 6617624A4F23CSAE CRC64;

Query Match 100.0%; Score 19; DB 1; Length 152;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 92 VAEF 95

RESULT 28
RIBB_SULSO
ID RIBB_SULSO STANDARD; PRT; 154 AA.
AC Q980B5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 6,7-dimethyl-8-ribityllumazine synthase (SC 2.5.1.9) (DMRL synthase)
DE (Lumazine synthase) (Riboflavin synthase beta chain).
EN RIBB OR SSC0400.
DS Sulfolobus solfataricus.
DC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
DC Sulfolobus.
DX NCBI_TaxID=2287;
ZN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
ZA Awavez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moers A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,

```

```

PA Garrett R.A., Ragan M.A., Sensen C.W., Van der Coost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
CC ribityl-amino-2,4(1H,3H)-pyrimidinone and L-3,4-dihydroxy-2-
CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-
CC amino-2,4(1H,3H)-pyrimidinone with L-3,4-dihydroxy-2-
CC butanone-4-phosphate yielding 6,7-dimethyl-8-lumazine (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1-D-ribityl)lumazine =
CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
CC -!- PATHWAY: Riboflavin biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the DMRL synthase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE006673; AAK40729.1; -.
CC PIR; B90184; B90184.
CC HAMAP; MF 00178; -. 1.
CC InterPro; IPR002180; DMRL synthase.
CC Pfam; PF00885; DMRL synthase; 1.
CC ProDom; PD003664; DMRL synthase; 1.
CC TIGRFAMs; TIGR00114; RibH; 1.
KW Riboflavin biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 154 AA; 17247 MW; ADFA9E6A2C723210 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 154;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 12 VAEF 15

RESULT 29
PCP_HAEIN
ID PCP_HAEIN STANDARD; PRT; 155 AA.
AC P10325;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Outer membrane lipoprotein PCP precursor (15 kDa lipoprotein) (PAL
DE cross-reacting lipoprotein).
DE PCP OR LPP OR H1579.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
CX NCBI_TaxID=727;
ZN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88115138; PubMed=2828309;
RA Deich R.A., Metcalf B.J., Finn C.W., Farley J.E., Green B.A.;
RT "Cloning of genes encoding a 15,000-dalton peptidoglycan-associated
RT outer membrane lipoprotein and an antigenically related 15,000-dalton
RT protein from Haemophilus influenzae.";
RL J. Bacteriol. 170:489-498(1988).
ZN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kierlavage A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

```

Weidman J.F., Phillips C.A., Spriggs T., Redblom E., Cotton M.D.,  
 Utterback T.R., Hanna M.C., Nguyen D.T., Saudex D.M., Brandon R.C.,  
 Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 Venter J.C.;  
 "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 Science 269:496-512(1995).  
 -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 anchor.  
 -!- SIMILARITY: TO E. COLI AND S. TYPHIMURIUM SLYS AND TO  
 Y. ENTEROCOLITICA PCP.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: M18977; AAC24938.1; -;  
 EMBL: U32832; AAC23228.1; -;  
 PIR: I64130; I64130.  
 TIGR: H11579; -;  
 InterPro: IPR000437; Prok lipoprot S.  
 PROSITE: PS00013; PROKAR LIPOPROTEIN; 1.  
 Outer membrane; Lipoprotein; Signal; Complete proteome; Palmitate.  
 SIGNAL 1 18  
 CHAIN 19 155 OUTER MEMBRANE LIPOPROTEIN PCP.  
 LIPID 19 19 N-palmitoyl cysteine.  
 LIPID 19 19 S-diacylglycerol cysteine.  
 CONFLICT 135 143 CSLVAEFVF -> VAGRRVRI (IN REF. 1).  
 SEQUENCE 155 AA; 15425 MW; D7880327FCFC985 CRC64;  
 -----  
 Query Match 100.0%; Score 19; DB 1; Length 155;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 VAEF 4  
 ||||  
 138 VAEF 141  
 -----  
 RESULT 30  
 RAP TAROF STANDARD; PRT; 157 AA.  
 049765;  
 30-MAY-2000 (Rel. 39, Created)  
 30-MAY-2000 (Rel. 39, Last sequence update)  
 15-MAR-2004 (Rel. 43, Last annotation update)  
 Root allergen protein (RAP).  
 Taraxacum officinale (Common dandelion).  
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae;  
 Taraxacum.  
 NCBI\_TaxID=50225;  
 [1]  
 SEQUENCE FROM N.A.  
 TISSUE:Root;  
 Xu X.-Y., Bewley J.D., Greenwood J.S.;  
 Submitted (DEC-1997) to the EMBL/GenBank/DBSJ databases.  
 -!- ALLERGEN: Causes an allergic reaction in human.  
 -!- SIMILARITY: Belongs to the BetVI family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

CC EMBL: AF036931; AAB92255.1; -;  
 DR HSRP; C24248; 1809.  
 DR InterPro: IPR000916; Bet\_v\_I.  
 DR Pfam: PF00407; Bet\_v\_I.  
 DR PRINTS: PR00634; BETALLERGEN.  
 DR PROSITE: PS00451; PATHOGENESIS BETVI; 1.  
 KW Allergen; Plant defense; Pathogenesis-related protein.  
 SQ SEQUENCE 157 AA; 17040 MW; 5892AB9593A8A7E0 CRC64;  
 -----  
 Query Match 100.0%; Score 19; DB 1; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 VAEF 4  
 ||||  
 3 VAEF 6  
 -----  
 RESULT 31  
 RISE SULTO STANDARD; PRT; 157 AA.  
 ID RISE SULTO STANDARD; PRT; 157 AA.  
 AC Q975M5;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 6,7-dimethyl-8-ribitylumazine synthase (EC 2.5.1.9) (DMRL synthase)  
 (Lumazine synthase) (Riboflavin synthase beta chain).  
 GN RIBH OR SP0394.  
 OS Sulfolobus tokodaii.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=111955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JCM 10545 / 7;  
 RA MEDLINE=21456156; PubMed=11572479;  
 RA Kawabayashi Y., Hiro Y., Horikawa H., Jin-no K., Takahashi M.,  
 Sekine M., Baba S.-I., Ankal A., Kosugi H., Hosoyama A., Fukui S.,  
 Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 Yoshizawa T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,  
 Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
 Oshima T., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic thermoacidophilic  
 Crenarchaeon, Sulfolobus tokodaii strain7.";  
 RL DNA Res. 8:123-140(2001).  
 CC -!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex  
 catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-  
 ribityl-amino-2,4(1H,3H)-pyrimidinedione and L-3,4-dihydroxy-2-  
 butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit  
 catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-  
 amino-2,4(1H,3H)-pyrimidinedione with L-3,4-dihydroxy-2-  
 butanone-4-phosphate yielding 6,7-dimethyl-8-lumazine (By  
 similarity).  
 CC -!- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1-D-ribityl)lumazine =  
 riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.  
 CC -!- PATHWAY: Riboflavin biosynthesis; last step.  
 CC -!- SIMILARITY: Belongs to the DMRL synthase family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: AF000982; BAB65375.1; -;  
 DR HAMAP; MF 00178; -; 1.  
 DR InterPro: IPR002180; DMRL synthase.  
 DR Pfam: PF00885; DMRL synthase; 1.  
 DR ProDom: PD003664; DMRL synthase; 1.  
 DR TIGRFAMs: TIGR00114; ribH; 1.  
 DR

```
KW Riboflavin biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 157 AA; 17501 MW; 529558077E164A18 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 157;
Best Local Similarity 100.0%; Pred. No. 2.9e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0;

2Y 1 VAEF 4
2b 14 VAEF 17

RESULT 32
SGSS_DROME STANDARD; PRT; 163 AA.
AC PG7701; Q9VE15;
JT 01-APR-1988 (Rel. 07, Created)
JT 01-APR-1988 (Rel. 07, Last sequence update)
JT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Salivary glue protein Sgs-5 precursor.
EN SGS5 OR CG7596.
XS Drosophila melanogaster (Fruit fly).
XC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
XC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
XC Ephydroidea; Drosophilidae; Drosophila.
XK NCBI_TaxID=7227;
XN [1]
XP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=87086754; PubMed=3099981;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RX Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RX George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RX Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RX Brandon R.C., Rogers Y.-H.C., Blazei R.G., Champe M., Pfeiffer B.D.,
RX Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.B.G.,
RX Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RX Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RX Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RX Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RX Burtis K.C., Busam M.A., Butler H., Cadieu E., Center A., Chandra I.,
RX Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RX de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RX Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RX Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RX Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RX Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RX Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RX Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RX Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RX Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RX Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RX Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RX Mervulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RX Mount S.M., Moy M., Murphy B., Murphy L., Musker D.R., Nelson D.L.,
RX Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,
RX Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RX Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,
RX Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Sun E.,
RX Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RX Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RX Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RX Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RX Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RX Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
```

```

R EMBL; L22173; AAA34936.1; -
R EMBL; S65964; AAD13970.1; -
R EMBL; S66120; AAB28442.1; -
R EMBL; U87779; AAB65007.1; -
R PIR; S50509; S50509.
R GerMOnline; 139039; -
R SGD; S0000761; UTR5.
T T CONFLICT 1 20
T MSRYGKLVHYIIVHDDQR -> MRDSNVKISVFPALYN
T RGNNTIN (IN REF. 1).
Q SEQUENCE 166 AA; 19335 MW; A58EE8D0FA6D3DEA CRC64;

Query Match 100.0%; Score 19; DB 1; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4
b 92 VAEF 95

ESULT 34
SB_RHOBA STANDARD; PRT; 169 AA.
C PS9332; Q7UKV3.
T 15-MAR-2004 (Rel. 43, Created)
T 15-MAR-2004 (Rel. 43, Last sequence update)
T 15-MAR-2004 (Rel. 43, Last annotation update)
E Single-strand binding protein (SSB) (Helix-destabilizing protein).
N SSB OR RB9917.
S Rhodopirellula baltica.
C Bacteria; Planctomycetales; Planctomycetacia; Planctomycetales;
C Planctomycetaceae; Firellula.
X NCBI_TaxID=117;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=1;
X MEDLINE=22735913; PubMed=12835416;
A Gloeckner F.O., Kube M., Bauer M., Tesling H., Lombardot T.,
A Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
A Schlessner H., Amann R., Reinhardt R.;
A "Complete genome sequence of the marine planctomycete Firellula sp.
T strain 1.";
T Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
C L - FUNCTION: This protein is essential for replication of the
C chromosome. It is also involved in DNA recombination and repair
C (by similarity).
C -1- SIMILARITY: Contains 1 SSB domain.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
C EMBL; BX294150; CAD76529.1; -
C PROSITE; PS0935; SSB; 1.
W DNA-binding; DNA repair; DNA replication; Complete proteome.
T DOMAIN 4 107
T DOMAIN 113 121 POLY-GLY.
Q SEQUENCE 169 AA; 18108 MW; AA8F93B0FA51F287 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 169;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4
b 28 VAEF 31

ESULT 35

```

```

BFL1 MOUSE STANDARD; PRT; 172 AA.
AC Q07440;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bcl-2-related protein A1 (BFL-1 protein) (hemopoietic-specific early
DE response protein) (A1-A).
GN BCL2A1 OR BCL2A1A OR BFL1 OR A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBA/J; TISSUE=Bone marrow;
RX MEDLINE=93346743; PubMed=8345191;
RA Lin E.Y., Orlofsky A., Berger M.S., Prystowsky M.B.;
RT "Characterization of A1, a novel hemopoietic-specific early-response
RT gene with sequence similarity to bcl-2.";
RL J. Immunol. 151:1979-1988(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=Liver;
RX MEDLINE=98307518; PubMed=9645611;
RA Hatakeyama S., Hamasaki A., Negishi I., Loh D.Y., Sendo F.,
RA Nakayama K., Nakayama K.-I.;
RT "Multiple gene duplication and expression of mouse bcl-2-related
RT genes, A1.";
RL Int. Immunol. 10:631-637(1998).
CC -1- FUNCTION: Retards apoptosis induced by IL-3 deprivation. May
CC function in the response of hemopoietic cells to external signals
CC and in maintaining endothelial survival during infection.
CC -1- SUBCELLULAR LOCATION: Intracellular.
CC -1- TISSUE SPECIFICITY: Expressed in hemopoietic tissues, including
CC bone marrow, spleen and thymus.
CC -1- INDUCTION: By granulocyte-macrophage colony-stimulating factor
CC and LPS in macrophages.
CC -1- SIMILARITY: Contains 1 Bcl-2 homology 1 (BH1) domain.
CC -1- SIMILARITY: Contains 1 Bcl-2 homology 2 (BH2) domain.
CC -1- SIMILARITY: Belongs to the Bcl-2 family.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
C EMBL; L16462; AAA16886.1; -
C EMBL; U23774; AAB97953.1; -
C EMBL; U23773; AAB97953.1; JOINED.
C PIR; I49449; I49449.
C HSP; Q07817; IMAZ.
C MGD; MGI:102687; Bcl2ala.
C InterPro; IPR000712; Bcl2_BH.
C InterPro; IPR002475; BCL2_family.
C Pfam; PF00452; Bcl-2; 1.
C SMART; SM00337; BCL; 1.
C PROSITE; PS0062; BCL2_FAMILY; 1.
C PROSITE; PS01090; BH1; 1.
C PROSITE; PS01258; BH2; 1.
C Apoptosis.
C DOMAIN 24 33 ALA/PRO-RICH.
FT DOMAIN 77 97 BH1.
FT DOMAIN 132 147 BH2.
SQ SEQUENCE 172 AA; 19914 MW; 37AD35818E756488 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 172;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 VAEF 4  
DB 122 VAEF 125

RESULT 36  
BCL2 HUMAN  
ID BCL2 HUMAN STANDARD; PRT; 175 AA.  
AC Q16548; Q99524; PRT; 175 AA.  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Bcl-2-related protein A1 (BCL-1 protein) (Hemopoietic-specific early response protein) (GRS protein).  
GN BCL2A1 OR BCL1 OR GRS OR BCL215.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Umbilical vein;  
RX MEDLINE=96184764; PubMed=8605321;  
RA Karson A., Yee E., Kaushansky K., Harlan J.M.;  
RT "Cloning of human Bcl-2 homologue: inflammatory cytokines induce human A1 in cultured endothelial cells."  
RL Blood 87:3089-3096(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=96068995; PubMed=7478596;  
RA Choi S.S., Park I.-C., Yun J.W., Sung Y.C., Hong S.-I., Shin H.-S.;  
RT "A novel Bcl-2 related gene, Bcl-1, is overexpressed in stomach cancer and preferentially expressed in bone marrow."  
RL Oncogene 11:1693-1698(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=T-cell;  
RX MEDLINE=97203281; PubMed=9050999;  
RA Kenny J.J., Knobloch T.J., Augustus M., Carter K.C., Rosen C.A., Lang J.C.;  
RT "GRS, a novel member of the Bcl-2 gene family, is highly expressed in multiple cancer cell lines and in normal leukocytes."  
RL Oncogene 14:997-1001(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.R., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S.S., Garcia A.M., Gay L.J., Rulyk S.W., Villalón D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -1- FUNCTION: Retards apoptosis induced by IL-3 deprivation. May function in the response of hemopoietic cells to external signals and in maintaining endothelial survival during infection (By similarity). LOCATION: Intracellular.

CC -1- TISSUE SPECIFICITY: Seems to be restricted to the hematopoietic compartment. Expressed in peripheral blood, spleen, and bone marrow, at moderate levels in lung, small intestine and testis, at a minimal levels in other tissues. Also found in vascular smooth muscle cells and hematopoietic malignancies.  
CC -1- INDUCTION: By phorbol ester and inflammatory cytokines, such as TNF-alpha, or IL-1-beta, but not by growth factors.  
CC -1- SIMILARITY: Contains 1 Bcl-2 homology 1 (BH1) domain.  
CC -1- SIMILARITY: Contains 1 Bcl-2 homology 2 (BH2) domain.  
CC -1- SIMILARITY: Belongs to the Bcl-2 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U29680; AAC50438.1; -  
CC EMBL; U27467; AAC50288.1; -  
CC EMBL; Y09397; CAA70566.1; -  
CC EMBL; BC016281; AAH16281.1; -  
CC PIR; I39055; I39055.  
CC HSP; P53563; IAE3.  
CC Genew; HGNC:991; BCL2A1.  
CC MIM; 601056; -  
CC GO; GO:0008189; P:apoptosis inhibitor activity; TAS.  
CC GO; GO:0006916; P:anti-apoptosis; TAS.  
CC InterPro; IPR000712; BCL2\_BH.  
CC InterPro; IPR002475; BCL2\_family.  
CC Pfam; PF00452; Bcl-2; 1.  
CC SMART; SM00337; BCL; 1.  
CC PROSITE; PS50062; BCL2\_FAMILY; 1.  
CC PROSITE; PS01080; BH1; 1.  
CC PROSITE; PS01258; BH2; 1.  
KW Apoptosis.  
FT DOMAIN 24 33 ALA/PRO-RICH.  
FT DOMAIN 77 97 BH1.  
FT DOMAIN 132 147 BH2.  
FT CONFLICT 72 72 N -> T (IN REF. 3).  
FT CONFLICT 107 107 Q -> H (IN REF. 3).  
SQ SEQUENCE 175 AA; 20132 MW; 329D98AF2BE07A0D CRC64;

Query Match 100.0%; Score 19; DB 1; Length 175;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 122 VAEF 125

RESULT 37  
FR12-RANCA  
ID FR12-RANCA STANDARD; PRT; 176 AA.  
AC P07798;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ferritin, middle subunit (Ferritin M) (Ferritin X) (Ferritin H').  
OS Rana catesbeiana (Bull frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Rana.  
NCBI\_TaxID=8400;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87222424; PubMed=3495534;  
RA Dickey L.F., Sreedharan S., Theil B.C., Digsbury J.R., Wang Y.-H., Kaufman R.E.;  
RT "Differences in the regulation of messenger RNA for housekeeping and specialized-cell ferritin. A comparison of three distinct ferritin complementary DNAs, the corresponding subunits, and identification of

the first processed in amphibia.";  
J. Biol. Chem. 262:7901-7907(1987).  
[2]  
X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
MEDLINE=99367924; PubMed=10439069;  
Ha Y., Shi D., Small G.W., Thell B.C., Allewell N.M.,  
"Crystal structure of bullfrog M ferritin at 2.8 A resolution:  
analysis of subunit interactions and the binuclear metal center.";  
J. Biol. Inorg. Chem. 4:243-256(1999).  
-!- FUNCTION: Ferritin is an intracellular molecule that stores iron  
in a soluble, nontoxic, readily available form. The functional  
molecule, which is composed of 24 chains, is roughly spherical and  
contains a central cavity into which the polymeric ferric iron  
core is deposited.  
-!- MISCELLANEOUS: THERE ARE THREE TYPES OF FERRITIN SUBUNITS: L, M  
AND H CHAINS IN AMPHIBIA.  
-!- SIMILARITY: Belongs to the ferritin family.  
-!- SIMILARITY: Contains 1 ferritin-like diiron domain.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; J02724; AAA49525.1; -;  
PIR; C27805; C27805.  
PDB; 1MPR; 22-JUN-99.  
InterPro; IPR001519; Ferritin.  
InterPro; IPR008331; Ferritin\_Ops.  
InterPro; IPR009040; Ferritin\_Like.  
Pfam; PF00210; ferritin; 1.  
ProDom; PD000971; Ferritin; 1.  
PROSITE; PS00540; FERRITIN\_1; 1.  
PROSITE; PS00204; FERRITIN\_2; 1.  
PROSITE; PS00905; FERRITIN\_LIKE; 1.  
Iron storage; Iron; Metal-Binding; 3D-structure.  
DOMAIN 7 156 FERRITIN-LIKE DIIRON.  
METAL 24 24 IRON (BY SIMILARITY).  
METAL 58 58 IRON (BY SIMILARITY).  
METAL 59 59 IRON (BY SIMILARITY).  
METAL 62 62 IRON (BY SIMILARITY).  
METAL 104 104 IRON (BY SIMILARITY).  
METAL 138 138 IRON (BY SIMILARITY).  
HELIX 11 39  
TURN 41 43  
HELIX 46 73  
TURN 74 74  
STRAND 82 82  
HELIX 93 120  
TURN 121 122  
HELIX 124 133  
TURN 134 134  
HELIX 135 154  
TURN 155 159  
HELIX 161 170  
TURN 171 171  
SEQUENCE 176 AA; A9P0F5EBE8584D46 CRC64;  
  
Query Match 100.0%; Score 19; DB 1; Length 176;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Y 1 VAEF 4  
b 48 VAEF 51  
|||||  
-----  
RESULT 38  
LI0\_THETN STANDARD; PRT; 177 AA.  
D LI0\_THETN

Q8RTU4;  
28-FEB-2003 (Rel. 41, Created)  
28-FEB-2003 (Rel. 41, Last sequence update)  
28-FEB-2003 (Rel. 41, Last annotation update)  
508 ribosomal protein L10.  
RPLJ OR TIE2304.  
GN Thermoanaerobacter tengcongensis.  
OS Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;  
OC Thermoanaerobacteriaceae; Thermoanaerobacter.  
NCBI\_TaxID=119072;  
OX [1]  
SEQUENCE FROM N.A.  
RP STRAIN=MB4 / JCM 11007;  
RC MEDLINE=21992816; PubMed=11977336;  
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
Chen Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
Tan H., Chen R., Wang J., Yu J., Yang H.;  
"A complete sequence of T. tengcongensis genome.";  
Genome Res. 12:689-700(2002)  
RL -!- SIMILARITY: Belongs to the L10P family of ribosomal proteins.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AE013173; AAM25445.1; -;  
DR HAMAP; MF 00362; -; 1.  
DR InterPro; IPR001790; Ribosomal\_L10.  
DR InterPro; IPR002363; Ribosomal\_L10eub.  
DR Pfam; PF00466; Ribosomal\_L10; 1.  
DR PROSITE; PS01109; RIBOSOMAL\_L10; FALSE\_NEG.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 177 AA; 19591 MW; 50DDCF896EF6F4E8 CRC64;  
  
Query Match 100.0%; Score 19; DB 1; Length 177;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VAEF 4  
Db 12 VAEF 15  
|||||  
-----  
RESULT 39  
DSBB\_PASMU STANDARD; PRT; 178 AA.  
ID DSBB\_PASMU  
AC Q9L6B3; P57804;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Disulfide bond formation protein B (Disulfide oxidoreductase).  
DS DSBB OR PM0046.  
GN Pasteurella multocida.  
OS Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Pasteurella.  
OX NCBI\_TaxID=747;  
RL [1]  
SEQUENCE FROM N.A.  
RP FULLER T.E., Kennedy M.J., Lowery D.E.;  
"Identification of Pasteurella multocida virulence genes in a  
septicemic mouse model using signature-tagged mutagenesis.";  
Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
SEQUENCE FROM N.A.  
RC STRAIN=Em70;  
RX MEDLINE=21145866; PubMed=11248100;  
RA May B.J., Zhang Q., Li L., Paustian M.L., Whittam T.S., Kapur V.;  
"Complete genomic sequence of Pasteurella multocida Em70.";  
Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

CC CC -!- FUNCTION: Required for disulfide bond formation in some  
 CC periplasmic proteins. Acts by oxidizing the dsba protein (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (By similarity).  
 CC -!- SIMILARITY: Belongs to the dsbB family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF237925; AAF68411.1; -;  
 CC EMBL: AE006038; AAK02130.1; -;  
 CC HAMAP: MF\_00286; -; 1.  
 CC InterPro: IPR003752; DsbB.  
 CC Pfam: PF02600; DsbB; 1.  
 CC Oxidoreductase; Redox-active center; Transport; Electron transport;  
 CC Chaperone; Transmembrane; Inner membrane; Complete proteome.  
 CC DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 15 31 POTENTIAL.  
 CC DOMAIN 32 49 PERIPLASMIC (POTENTIAL).  
 CC TRANSMEM 50 65 POTENTIAL.  
 CC DOMAIN 66 72 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 73 90 POTENTIAL.  
 CC DOMAIN 91 145 PERIPLASMIC (POTENTIAL).  
 CC TRANSMEM 146 164 POTENTIAL.  
 CC DOMAIN 165 177 CYTOPLASMIC (POTENTIAL).  
 CC DISULFID 41 44 REDOX-ACTIVE (BY SIMILARITY).  
 CC DISULFID 105 131 REDOX-ACTIVE (BY SIMILARITY).  
 CC SEQUENCE 178 AA; 20036 MW; D2C45BB73B31F0BC CRC64;

Query Match 100.0%; Score 19; DB 1; Length 178;

Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dy 1 VAEF 4

Db 108 VAEF 111

RESULT 40

ESM5 DROME

ID ESM5 DROME STANDARD; PRT; 178 AA.

PI3096; Q9VB18;

01-JAN-1990 (Rel. 13, Created)

01-JAN-1990 (Rel. 13, Last sequence update)

15-VAR-2004 (Rel. 43, Last annotation update)

Enhancer of split m5 protein (E(spl)m5).

HLHMS OR CG6096.

Drosophila melanogaster (Fruit fly).

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

NCBI\_TaxID=7227;

SEQUENCE FROM N.A.

MEDLINE=89231619; PubMed=2540957;

Klaembt C., Knust E., Tietze K., Campos-Ortega J.A.;

"Closely related transcripts encoded by the neurogenic gene complex

enhancer of split of Drosophila melanogaster.";

EMBO J. 8:203-210(1989).

SEQUENCE FROM N.A.

STRAIN=Berkeley;

MEDLINE=20196006; PubMed=10731132;

Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,

Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Balwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Saunders R.D.C., Scheeler F., Shen H.,  
 Reinert K., Remington K., Simpson M., Skupski M.P., Smith T.,  
 Shue B.C., Siden-Kiamos I., Stapleton M., Strong R., Sun E.,  
 Spier E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,  
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhu S., Zhu X., Smith H.O.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.).  
 "The genome sequence of Drosophila melanogaster.";  
 Science 287:2185-2195(2000).  
 [3]

WRPW MOTIF.

MEDLINE=95094252; PubMed=8001118;

Paroush Z., Finley R.B. Jr., Kidd T., Wainwright S.M., Ingham P.W.,

Brent R., Ish-Horowicz D.;

"Groucho is required for Drosophila neurogenesis, segmentation, and

sex determination and interacts directly with hairy-related bHLH

proteins.";

Cell 79:805-815(1994).

-!- FUNCTION: Participates in the control of cell fate choice by

uncommitted neuroectodermal cells in the embryo. Transcriptional

repressor. Binds DNA on N-box motifs: 5'-CACNAG-3'.

-!- SUBUNIT: Transcription repression requires formation of a complex

with a co-repressor protein (Groucho). Forms homodimers.

-!- SUBCELLULAR LOCATION: Nuclear (Probable).

-!- DEVELOPMENTAL STAGE: Expressed at the time when separation of

neural and epidermal precursors cells occurs. Mesectodermal

expression appears shortly before the onset of gastrulation.

-!- DOMAIN: The orange domain and the basic helix-loop-helix motif

mediate repression of specific transcriptional activators, such

as basic helix-loop-helix protein dimers.

-!- DOMAIN: The C-terminal WRPW motif is a transcriptional repression

domain necessary for the interaction with Groucho, a

transcriptional co-repressor recruited to specific target DNA by

Hairy-related proteins.

-!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.

-!- SIMILARITY: Contains 1 orange domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL: X16552; CAA34552.1; -;  
 EMBL: AE003754; AAF56552.1; -;  
 PIR: S03629; S03629.

```
3 TRANSFAC; T01644; -.
3 FlyBase; FBgn0002631; HLMs.
3 GO; GO:0005634; C:nucleus; IDA.
3 GO; GO:0003677; C:DNA binding; IDA.
3 InterPro; IPR001092; HLH basic.
3 InterPro; IPR003650; Orange.
3 Pfam; PF00010; HLH; 1.
3 SMART; SM00353; HLH; 1.
3 SMART; SM00511; ORANGE; 1.
3 PROSITE; PS00888; HLH; 1.
3 Differentiation; Neurogenesis; Nuclear protein; DNA-binding;
3 Transcription regulation; Repressor.
3 DNA BIND 19 33 BASIC DOMAIN.
3 DOMAIN 34 74 HELIX-LOOP-HELIX MOTIF.
3 DOMAIN 86 129 ORANGE.
3 DOMAIN 175 178 WRPW MOTIF.
3 SEQUENCE 178 AA; 19923 MW; 19363D0F6043C84F CRC64;
Query Match 100.0%; Score 19; DB 1; Length 178;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 VAEF 4
D 45 VAEF 48
RESULT 41
B18_ARCFU STANDARD; PRT; 178 AA.
D YB18_ARCFU
C 029147;
I 16-OCT-2001 (Rel. 40, Created)
I 16-OCT-2001 (Rel. 40, Last sequence update)
I 16-OCT-2001 (Rel. 40, Last annotation update)
3 Hypothetical protein AF1118 precursor.
3 AF1118.
3 Archaeoglobus fulgidus.
3 Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
3 Archaeoglobaceae; Archaeoglobus.
3 NCBI_TaxID=2234;
[1]
N SEQUENCE FROM N.A.
P STRAIN=VC-16 / DSM 4304 / ATCC 49558;
X MEDLINE=98049343; PubMed=9389475;
A Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
A Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
A Richardson D.L., Kariavage A.R., Graham D.E., Kyriades N.C.,
A Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
A Kirkness E.P., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
A Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
A Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
A Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
A Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
A Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
A Venter J.C.;
I "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus."
L Nature 390:364-370(1997).
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
C -----
R EMBL; AE001027; AAB90141.1; -.
R PIR; E69389; E69389.
R TIGR; AF1118; -.
W Hypothetical protein; Signal; Complete proteome.
T SIGNAL 1 20 POTENTIAL.
T CHAIN 21 178 HYPOTHETICAL PROTEIN AF1118.
SQ SEQUENCE 178 AA; 19319 MW; 378A4F200240D924 CRC64;
Query Match 100.0%; Score 19; DB 1; Length 178;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAEF 4
DB 155 VAEF 158
RESULT 42
APT_HAEDU STANDARD; PRT; 179 AA.
AC Q7VZQ4;
AT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).
GN APT OR HD1818.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OC NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=35000HP / ATCC 700724;
RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT "The complete genome sequence of Haemophilus ducreyi."
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Catalyzes a salvage reaction resulting in the formation
CC of AMP, that is energetically less costly than de novo synthesis.
CC -!- CATALYTIC ACTIVITY: AMP + diphosphate = adenine + 5-phospho-alpha-
CC D-ribose 1-diphosphate.
CC -!- PATHWAY: Purine salvage.
CC -!- SUBUNIT: Homodimer (by similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the purine/pyrimidine
CC phosphoribosyltransferase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB017156; AAP96568.1; -.
DR HAMAP; MF_000084; -.
DR InterPro; IPR005764; Ado_phospho_trans.
DR InterPro; IPR002375; Pr/py_rp_transf.
DR InterPro; IPR000836; PRTtransferase.
DR Pfam; PF00156; Pribosyltran; 1.
DR TIGRfam; TIGR01090; apt; 1.
DR PROSITE; PS00103; PUR_PYR_PR_TRANSFER; 1.
KW transferase; Glycosyltransferase; Purine salvage; Complete proteome.
SQ SEQUENCE 179 AA; 19315 MW; 13A1DAFDC13560DF CRC64;
Query Match 100.0%; Score 19; DB 1; Length 179;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAEF 4
DB 45 VAEF 48
RESULT 43
RETE_BOVIN STANDARD; PRT; 183 AA.
ID RETE_BOVIN
AC P18902;
SQ SEQUENCE 179 AA; 19315 MW; 13A1DAFDC13560DF CRC64;
Query Match 100.0%; Score 19; DB 1; Length 179;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Plasma retinol-binding protein (PRBP) (RBP).  
GN RBP4.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
[1]  
RN SEQUENCE.  
RA MEDLINE=91006139; PubMed=2209607;  
RX Zanotti G.; Berni R.; Stoppani M.; Zapponi M.C.; Meloni M.L.; Monaco H.L.;  
RA "The bovine plasma retinol-binding protein. Amino acid sequence,  
RT interaction with transthyretin, crystallization and preliminary X-ray  
RT data.";  
RL Eur. J. Biochem. 192:507-513(1990).  
[2]  
RN SEQUENCE OF 52-183 FROM N.A.  
RX MEDLINE=93385352; PubMed=8373966;  
RA Liu K.H.; Dore J.J. Jr.; Roberts M.P.; Krishnan R.; Hopkins F.M.;  
RA Godkin J.D.;  
RT "Expression and cellular localization of retinol-binding protein  
RT messenger ribonucleic acid in bovine blastocysts and extraembryonic  
RT membranes.";  
RL Biol. Reprod. 49:393-400(1993).  
[3]  
RN X-RAY CRYSTALLOGRAPHY.  
RX MEDLINE=92322903; PubMed=1623143;  
RA Monaco H.L.; Zanotti G.;  
RT "Three-dimensional structure and active site of three hydrophobic  
RT molecule-binding proteins with significant amino acid sequence  
RT similarity.";  
RL Biopolymers 32:457-465(1992).  
[4]  
RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RX MEDLINE=93266508; PubMed=8496140;  
RA Zanotti G.; Berni R.; Monaco H.L.;  
RT "Crystal structure of liganded and unliganded forms of bovine plasma  
RT retinol-binding protein.";  
RL J. Biol. Chem. 268:10728-10738(1993).  
CC -!- FUNCTION: Delivers retinol from the liver stores to the peripheral  
CC tissues. In plasma, the RBP-retinol complex interacts with  
CC transthyretin, this prevents its loss by filtration through the  
CC kidney glomeruli.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: Belongs to the lipocalin family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; S65585; AAB28336.1; -.  
DR PIR; I46955; I46955.  
DR PIR; S13186; S13186.  
DR PDB; 1REB; 31-JAN-94.  
DR PDB; 1HBP; 31-JAN-94.  
DR PDB; 1HBQ; 31-JAN-94.  
DR PDB; 1FEL; 01-NOV-94.  
DR PDB; 1FEM; 01-NOV-94.  
DR PDB; 1FEN; 01-NOV-94.  
DR InterPro; IPR002345; Lipocalin.  
DR InterPro; IPR000566; Lipocalin\_cytrBP.  
DR Pfam; PF00061; Lipocalin; 1.  
DR PRINTS; PR00179; LIPOCALIN.  
DR PROSITE; PS00213; LIPOCALIN; 1.  
KW Plasma; Vitamin A; Retinol-binding; Transport; Liver; Lipocalin;

KW 3D-structure. 4 160 BY SIMILARITY.  
FT DISULFID 70 174 BY SIMILARITY.  
FT DISULFID 120 129 BY SIMILARITY.  
FT HELIX 6 8  
FT TURN 13 14  
FT TURN 17 20  
FT STRAND 22 30  
FT TURN 33 34  
FT STRAND 39 47  
FT TURN 49 50  
FT STRAND 53 62  
FT TURN 64 65  
FT STRAND 68 79  
FT TURN 83 84  
FT STRAND 85 92  
FT TURN 95 96  
FT STRAND 100 109  
FT STRAND 114 123  
FT TURN 125 126  
FT STRAND 129 138  
FT TURN 141 142  
FT HELIX 146 158  
FT TURN 159 160  
FT STRAND 162 163  
FT TURN 166 167  
FT TURN 173 175  
SQ SEQUENCE 183 AA; 21068 MW; D6BA064CB9567C09 CRC64;  
Query Match 100.0%; Score 19; DB 1; Length 183;  
Best Local Similarity 100.0%; Pred. No. 3.2e-02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
|||  
DB 42 VAEF 45  
RESULT 44  
ID DNAA WOLSP STANDARD; PRT; 186 AA.  
AC P35907;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Chromosomal replication initiator protein dnaA (Fragment).  
GN DNAA.  
OS Wolbachia sp.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Wolbachiae; Wolbachia.  
OX NCBI\_TaxID=956;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9520115; PubMed=7894745;  
RA Bourdis K.; Nigianaki A.; Onyango P.; Savakis C.;  
RT "A prokaryotic dnaA sequence in *Drosophila melanogaster*: Wolbachia  
RT infection and cytoplasmic incompatibility among laboratory strains.";  
RL Insect Mol. Biol. 3:131-142(1994).  
CC -!- FUNCTION: Plays an important role in the initiation and regulation  
CC of chromosomal replication. Binds to the origin of replication; it  
CC binds specifically double-stranded DNA at a 9 bp consensus (dnaA  
CC box): 5'-TTATC(C/A)(C/A)-3'. DnaA binds to ATP and to acidic  
CC phospholipids (By similarity).  
CC -!- SIMILARITY: Belongs to the dnaA family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

```

R EMBL; Z28981; CAAB2285.1; -.
R PIR; S39317; S39317.
R HAMAP; MF_00377; -.
R InterPro; IPR001957; Bac_DnaA.
R Pfam; PF00308; bac_dnaA; 1.
R PRINTS; PRO0051; DNAA.
R PROSITE; PS01008; DNAA; PARTIAL.
R DNA replication; DNA-binding; ATP-binding.
I NON_TER 1
I TER 186
I SEQUENCE 186 AA; 21283 MW; C423C06CEFFB0459 CRC64;
Query Match 100.0%; Score 19; DB 1; Length 186;
Best Local Similarity 100.0%; Pred.No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 VAEF 4
b 134 VAEF 137

RESULT 45
CBI PARDE STANDARD; PRT; 190 AA.
C POS417;
I 01-NOV-1988 (Rel. 09, Created)
I 01-NOV-1988 (Rel. 09, Last sequence update)
I 15-MAR-2004 (Rel. 43, Last annotation update)
E Ubiquinol-cytochrome c reductase iron-sulfur subunit (EC 1.10.2.2)
E (Rieske iron-sulfur protein) (RISP).
E PETA.
S Paracoccus denitrificans.
C Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
C Rhodobacteraceae; Paracoccus.
X NCBI_TaxID=266;
N [1]
P SEQUENCE FROM N.A.
X MEDLINE=88007612; PubMed=2820981;
X Kurowski B., Ludwig B.;
I "The genes of the Paracoccus denitrificans bcl complex. Nucleotide
I sequence and homologues between bacterial and mitochondrial
I subunits."
J. Biol. Chem. 262:13805-13811(1987).
C -I- FUNCTION: Component of the ubiquinol-cytochrome c reductase
C complex (complex III or cytochrome b-c1 complex), which is a
C respiratory chain that generates an electrochemical potential
C coupled to ATP synthesis.
C -I- CATALYTIC ACTIVITY: QH(2) + 2 ferricytochrome c = Q + 2
C ferrocycytochrome c.
C -I- COFACTOR: Binds 1 2Fe-2S iron sulfur cluster per subunit (By
C similarity).
C -I- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
C cytochrome c1 and the Rieske protein.
C -I- SUBCELLULAR LOCATION: Membrane-bound.
C -I- MISCELLANEOUS: The Rieske protein is a high potential 2Fe-2S
C protein.
C -I- SIMILARITY: Belongs to the Rieske family.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C
C EMBL; M17522; AAA25571.1; -.
C EMBL; X05799; CAA29243.1; -.
C PIR; A29413; A29413.
C HGSP; P13272; IRIE.
R InterPro; IPR005805; Rieske.
R InterPro; IPR005805; Rieske.dom.
R InterPro; IPR006317; Rieske_proteo.

```

```

DR InterPro; IPR006311; Tat.
DR Pfam; PF00355; Rieske; 1.
DR PRINTS; PRO0162; RIESKE.
DR TIGRFAMs; TIGR01416; Rieske_proteo; 1.
DR TIGRFAMs; TIGR01409; Tat_signal_seq; 1.
DR PROSITE; PS00199; RIESKE_1; 1.
DR PROSITE; PS00200; RIESKE_2; 1.
KW Electron transport; Inner membrane; Transmembrane; Metal-binding;
KW Iron-sulfur; Iron; 2Fe-2S; Oxidoreductase.
FT TRANSMEM 18 39 POTENTIAL.
FT METAL 132 132 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 134 134 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 152 152 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 155 155 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT DISULFID 137 154 BY SIMILARITY.
SQ SEQUENCE 190 AA; 20299 MW; F83F5D9A3C1FBE84 CRC64;
Query Match 100.0%; Score 19; DB 1; Length 190;
Best Local Similarity 100.0%; Pred.No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAEF 4
Db 178 VAEF 181

RESULT 46
NTPA_METJA STANDARD; PRT; 193 AA.
AC Q57679;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Nucleoside-triphosphatase (EC 3.6.1.15) (Nucleoside triphosphate
DE phosphohydrolase) (NTPase).
GN NJ0226.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Karlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii";
RL Science 273:1058-1073(1996).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=99332061; PubMed=10404228;
RA Hwang K.Y., Chung J.H., Kim S.-H., Han Y.-S., Cho Y.;
RT "Structure-based identification of a novel NTPase from Methanococcus
RT jannaschii";
RL Nat. Struct. Biol. 6:691-696(1999).
CC -I- FUNCTION: CAN EFFICIENTLY HYDROLYZE NONSTANDARD NUCLEOTIDES SUCH
CC AS ATP TO XMP OR ITP TO IMP, BUT NOT THE STANDARD NUCLEOTIDES. XTP
CC IS THE BEST SUBSTRATE.
CC -I- CATALYTIC ACTIVITY: NTP + H(2)O = NDP + phosphate.
CC -I- COFACTOR: Magnesium or manganese.
CC -I- SUBUNIT: Homodimer.
CC -I- SIMILARITY: Belongs to the HAM1 NTPase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL; U67478; AAB98211.1; -  
 DR PIR; C64328; C64328.  
 DR PDB; 1B78; 28-JAN-00.  
 DR PDB; 2MJP; 28-JAN-00.  
 DR TIGR; M02226; -  
 DR HAMAP; MF01405; atypical; 1.  
 DR InterPro; IPR002637; Ham1p like.  
 DR Pfam; PF01725; Ham1p like; 1.  
 DR TIGRfam; TIGR00042; TIGR00042; 1.  
 KW Hydrolyase; Manganese; Magnesium; 3D-structure; Complete proteome.  
 FT STRAND 11 14  
 FT HELIX 18 27  
 FT TURN 28 30  
 FT TURN 32 33  
 FT TURN 36 39  
 FT STRAND 46 46  
 FT STRAND 50 65  
 FT HELIX 59 78  
 FT STRAND 69 78  
 FT HELIX 79 81  
 FT TURN 82 83  
 FT STRAND 85 86  
 FT TURN 87 88  
 FT HELIX 89 95  
 FT TURN 96 96  
 FT HELIX 97 105  
 FT TURN 106 107  
 FT STRAND 112 123  
 FT TURN 124 125  
 FT STRAND 126 138  
 FT HELIX 150 153  
 FT STRAND 155 157  
 FT TURN 158 159  
 FT HELIX 164 166  
 FT HELIX 169 172  
 FT TURN 173 175  
 FT HELIX 177 191  
 FT TURN 191 191  
 SQ SEQUENCE 193 AA; 22202 MW; 3570565B007D3DAC CRC64;

Query Match 100.0%; Score 19; DB 1; Length 193;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 DB 53 VAEF 56

RESULT 47  
 HM2\_BACHD  
 ID HAM2\_BACHD STANDARD; PRT; 194 AA.  
 AC G9X8D9;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE HAM1 protein homolog 2.  
 GN BH3067.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20312582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;

RT "Complete genome sequence of the alkaliphilic bacterium *Bacillus*  
 RT *halodurans* and genomic sequence comparison with *Bacillus subtilis*."  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 CC -!- SIMILARITY: Belongs to the HAM1 NTPase family.  
 CC -----

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL; AP001517; BAB06786.1; -  
 DR PIR; C84033; C84033.  
 DR HSPF; Q57679; 1B78.  
 DR HAMAP; MF\_01405; -; 1.  
 DR InterPro; IPR002637; Ham1p like.  
 DR Pfam; PF01725; Ham1p like; 1.  
 DR ProDom; PD004952; Ham1p like; 1.  
 DR TIGRfam; TIGR00042; TIGR00042; 1.  
 KW Hydrolase; Complete proteome.  
 SQ SEQUENCE 194 AA; 21868 MW; BECSA0C4F19A04B3 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 194;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 DB 14 VAEF 17

RESULT 48  
 TN90\_ECOLI  
 ID TN90\_ECOLI STANDARD; PRT; 194 AA.  
 AC P05823;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Transposon Tn2501 resolvase.  
 GN TNPR.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87103049; PubMed=3027041;  
 RA Michiels F., Cornelis G., Ellis K., Grinstead J.,  
 RT "Tn2501, a component of the lactose transposon Tn951, is an example  
 RT of a new category of class II transposable elements."  
 RL J. Bacteriol. 169:624-631(1987).  
 CC -!- FUNCTION: RESOLVASE CATALYZES THE RESOLUTION (A SITE-SPECIFIC  
 CC RECOMBINATION) OF THE COINTEGRATED REPLICON TO YIELD THE FINAL  
 CC TRANSPOSITION PRODUCTS.  
 CC -!- MISCELLANEOUS: TN2501 IS A CRYPTIC CLASS II TRANSPOSON FOUND AS  
 CC PART OF THE LACTOSE TRANSPOSON TN951.  
 CC -!- SIMILARITY: Belongs to the site-specific recombinase resolvase  
 CC family.  
 CC -----

This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL; M15197; AAA27426.1; -  
 DR PIR; B27758; RPRCR5.  
 DR HSPF; P03012; 2RSL.  
 DR InterPro; IPR006120; HTH\_7.

```

R InterPro: IPR006118; Recombinase.
R InterPro: IPR006119; resolvase_N.
R Pfam: PF02796; HTH_7; 1.
R Pfam: PF00239; resolvase; 1.
R PROSITE: PS00397; RECOMBINASES_1; 1.
R PROSITE: PS00398; RECOMBINASES_2; 1.
R DNA recombination; DNA integration; DNA-binding; Transposable element.
R ACT_SITE 11 11 TRANSIENT COVALENT LINKAGE TO DNA DURING
T STRAND CLEAVAGE AND REJOINING
T ACT_SITE 11 11 (BY SIMILARITY).
T I DNA BIND 170 189 H-T-H MOTIF (PROBABLE).
Q SEQUENCE 194 AA; 21420 MW; DEDFCBA6406EAD1 CRC64;
Query Match 100.0%; Score 19; DB 1; Length 194;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 VAEF 4
D 120 VAEF 123
RESULT 49
INQ_ECOLI STANDARD; PRT; 196 AA.
C P17170;
I 16-OCT-2001 (Rel. 40, Created)
I 16-OCT-2001 (Rel. 40, Last sequence update)
I 28-FEB-2003 (Rel. 41, Last annotation update)
I Putative DNA-invertase from lambdaoid prophage Qin.
X P17170 OR B1545.
S Escherichia coli.
S Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
S Enterobacteriaceae; Escherichia.
X NCBI_TaxID=562;
X [1]
X SEQUENCE FROM N.A.
X STRAIN=K12 / MG1655;
X MEDLINE=97426617; PubMed=9278503;
A Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
A Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
A Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
A Mau B., Shao Y.;
A "The complete genome sequence of Escherichia coli K-12.";
I Science 277:1453-1474 (1997).
X [2]
X SEQUENCE FROM N.A.
X STRAIN=K12;
X MEDLINE=97251357; PubMed=9097039;
A Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
A Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
A Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
A Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
A Sampaio G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
A Takenoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
A "A 570-kb DNA sequence of the Escherichia coli K-12 genome
I corresponding to the 28.0-40.1 min region on the linkage map.";
I DNA Res. 3:363-377 (1996).
X -!- SIMILARITY: Belongs to the site-specific recombinase resolvase
family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcements
or send an email to license@isb-sib.ch).
-----
EMBL; AE000252; AAC74618.1; -
EMBL; D90788; BAAJ5249.1; -
FIR; D64909; D64909.
HSSP; P03012; 2RSL.

```

```

DR EcoGene: EGI3824; pinQ.
DR InterPro: IPR006120; HTH_7.
DR InterPro: IPR006118; Recombinase.
DR InterPro: IPR006119; resolvase_N.
DR Pfam: PF02796; HTH_7; 1.
DR Pfam: PF00239; resolvase; 1.
DR PROSITE: PS00397; RECOMBINASES_1; FALSE_NEG.
DR PROSITE: PS00398; RECOMBINASES_2; 1.
KW Hypothetical protein; DNA recombination; DNA integration; DNA-binding;
KW DNA invertase; Complete proteome.
FT ACT_SITE 11 11 TRANSIENT COVALENT LINKAGE TO DNA DURING
FT STRAND CLEAVAGE AND REJOINING
FT (BY SIMILARITY).
SQ SEQUENCE 196 AA; 21852 MW; 7CF47F6D88BD497 CRC64;
Query Match 100.0%; Score 19; DB 1; Length 196;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAEF 4
DB 120 VAEF 123
RESULT 50
PINR_ECOLI STANDARD; PRT; 196 AA.
ID PINR_ECOLI STANDARD; PRT; 196 AA.
AC P77574;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative DNA-invertase from lambdaoid prophage Rac.
GN PINR OR B1374 OR C3146.
OS Escherichia coli, and
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992;
X [1]
X SEQUENCE FROM N.A.
X STRAIN=K12 / MG1655;
X MEDLINE=97426617; PubMed=9278503;
A Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
A Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
A Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
A Mau B., Shao Y.;
A "The complete genome sequence of Escherichia coli K-12.";
I Science 277:1453-1474 (1997).
X [2]
X SEQUENCE FROM N.A.
X STRAIN=K12;
X MEDLINE=97251357; PubMed=9097039;
A Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
A Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
A Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
A Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
A Sampaio G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
A Takenoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
A "A 570-kb DNA sequence of the Escherichia coli K-12 genome
I corresponding to the 28.0-40.1 min region on the linkage map.";
I DNA Res. 3:363-377 (1996).
X [3]
X SEQUENCE FROM N.A.
X STRAIN=O6:H1 / CFT073 / ATCC 700928;
X MEDLINE=22388234; PubMed=12471157;
A Welch R.A., Burland V., Plunkett G. III, Redford P., Roessch P.,
A Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
A Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
A Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
A "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
X -!- SIMILARITY: Belongs to the site-specific recombinase resolvase

```

```

CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000234; AAC74456.1; -
CC EMBL; D90775; BAA14579.1; -
CC EMBL; AE016764; AAN81596.1; -
CC PIR; A64888; A64888.
CC HSP; P03012; 2RSL.
CC EcoGene; EG13372; pinR.
CC InterPro; IPR006120; HTH_7.
CC InterPro; IPR006118; Recombinase.
CC InterPro; IPR006119; resolvase_N.
CC Pfam; PF02796; HTH_7; 1.
CC Pfam; PF0239; resolvase; 1.
CC PROSITE; PS00397; RECOMBINASES_1; FALSE_NEG.
CC PROSITE; PS00398; RECOMBINASES_2; 1.
CC Hypothetical protein; DNA recombination; DNA integration; DNA-binding;
CC DNA invertase; Complete proteome.
CC ACT_SITE 11
CC TRANSIENT COVALENT LINKAGE TO DNA DURING
CC STRAND CLEAVAGE AND REJOINING
CC (BY SIMILARITY).
CC SEQUENCE 196 AA; 21908 MW; 643A62BC940B29A9 CRC64;
CC
CC Query Match 100.0%; Score 19; DB 1; Length 196;
CC Best Local Similarity 100.0%; Pred. No. 3.4e+02;
CC Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC 1 VAEF 4
CC ||||
CC 120 VAEF 123
CC
CC RESULT 51
CC JHPA_ECOLI
CC ID _UHPA_ECOLI STANDARD; PRT; 196 AA.
CC AC P10940;
CC DT 01-JUL-1989 (Rel. 11, Created)
CC DT 01-JUL-1989 (Rel. 11, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Transcriptional regulatory protein uhpA.
CC GN UHPA OR B3669 OR C4593 OR Z5159 OR ECS4606 OR SF3792 OR S3976.
CC OS Escherichia coli.
CC OS Escherichia coli O6.
CC OS Escherichia coli O157:H7, and
CC Shigella flexneri.
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Escherichia.
CC NCBI_TaxID=562, 217992, 83334, 623;
CC [1] _SEQUENCE FROM N.A.
CC RC SPECIES=E.coli;
CC RA MEDLINE=87279903; PubMed=3301805;
CC RA Friedrich M.J., Kadner R.J.;
CC RT "Nucleotide sequence of the uhp region of Escherichia coli.";
CC RL J. Bacteriol. 169:3556-3563(1987).
CC [2]
CC RC SEQUENCE FROM N.A.
CC RC SPECIES=E.coli;
CC RX MEDLINE=92234930; PubMed=1569007;
CC RA Island M.D., Wei B.-Y., Kadner R.J.;
CC RT "Structure and function of the uhp genes for the sugar phosphate
CC transport system in Escherichia coli and Salmonella typhimurium.";
CC RL J. Bacteriol. 174:2754-2762(1992).
CC [3]
CC RC SEQUENCE FROM N.A.
CC RC SPECIES=E.coli; STRAIN=K12 / MG1655;

```

```

RX MEDLINE=93315143; PubMed=768682;
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RL genome: organizational symmetry around the origin of replication.";
RN Genomics 16:551-561(1993).
CC [4]
CC SEQUENCE FROM N.A.
CC RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
CC RX MEDLINE=22388234; PubMed=12471157;
CC RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
CC RA Rasko D., Buckles E.J., Liou S.-R., Boutin A., Hackett J., Stroud D.,
CC RA Mayhew G.F., Rose D.L., Zhou S., Schwartz D.C., Perna N.T.,
CC RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
CC RT "Extensive mosaic structure revealed by the complete genome sequence
CC of uropathogenic Escherichia coli.";
CC RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC [5]
CC SEQUENCE FROM N.A.
CC RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
CC RX MEDLINE=21074935; PubMed=11206551;
CC RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
CC RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
CC RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
CC RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
CC RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
CC RA Welch R.A., Blattner F.R.;
CC RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
CC RL Nature 409:529-533(2001).
CC [6]
CC SEQUENCE FROM N.A.
CC RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;
CC RX MEDLINE=21156231; PubMed=11258796;
CC RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
CC RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
CC RA Iida T., Takami H., Honda T., Sasaki G., Ogasawara N., Yasunaga T.,
CC RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
CC RT "Complete genome sequence of enterohemorrhagic Escherichia coli
CC O157:H7 and genomic comparison with a laboratory strain K-12.";
CC RL DNA Res. 8:11-22(2001).
CC [7]
CC SEQUENCE FROM N.A.
CC RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
CC RX MEDLINE=22272406; PubMed=12384590;
CC RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
CC RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
CC RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
CC RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
CC Yu J.;
CC RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
CC through comparison with genomes of Escherichia coli K12 and O157.";
CC RL Nucleic Acids Res. 30:4432-4441(2002).
CC [8]
CC SEQUENCE FROM N.A.
CC RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
CC RX MEDLINE=22590274; PubMed=12704152;
CC RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
CC RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
CC RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
CC RA Schwartz D.C., Blattner F.R.;
CC RT "Complete genome sequence and comparative genomics of Shigella
CC flexneri serotype 2a strain 2457T.";
CC RL Infect. Immun. 71:2775-2786(2003).
CC [9]
CC FUNCTION: APPEARS TO BE THE POSITIVE ACTIVATOR OF UHP
CC TRANSCRIPTION.
CC [10]
CC SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC [11]
CC PTM: PHOSPHORYLATED BY UHPB (PROBABLE).
CC [12]
CC SIMILARITY: Contains 1 response regulatory domain.
CC [13]
CC SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M17102; AAA24720.1; -  
EMBL; M89479; AAA24724.1; -  
EMBL; L10328; AAG62021.1; -  
EMBL; AE000444; AAC76592.1; -  
EMBL; AE018769; AAN83027.1; -  
EMBL; AE005596; AAG58868.1; -  
EMBL; AP002566; BAB38029.1; -  
EMBL; AE015386; AAN45232.1; ALT\_INIT.  
EMBL; AE016991; AAP18964.1; -  
PIR; A26925; BVECAU.  
PIR; F91204; F91204.  
PIR; H86050; H86050.  
HSSP; P10957; IRNL.  
ECODBASE; G020.9; 6TH EDITION.  
EcoGene; EG11051; uhpA.  
InterPro; IPR000792; HTH LuxR.  
InterPro; IPR001789; Response\_reg.  
Pfam; PF00196; GcrE; 1.  
Pfam; PF00072; response\_reg; 1.  
PRINTS; PR00038; HTHLUXR.  
ProDom; PD000307; HTH LuxR; 1.  
ProDom; PD000039; Response\_reg; 1.  
SMART; SM00421; HTH LUXR; 1.  
SMART; SM00448; REC\_1.  
PROSITE; PS00622; HTH LUXR FAMILY; 1.  
PROSITE; PS0110; RESPONSE REGULATORY; 1.  
Sensory transduction; Phosphorylation; Transcription regulation;  
DNA-binding; Activator; Sugar transport; Complete proteome.  
DOMAIN 1 116 RESPONSE REGULATORY.  
MOD RES 54 54 PHOSPHORYLATION (BY SIMILARITY).  
DNA\_BIND 155 174 H-T-H MOTIF (BY SIMILARITY).  
SEQUENCE 196 AA; 20889 MW; 8F409F8AADE08276 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 196;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
D 30 VAEF 33

RESULT 52  
UHPA\_SALTY STANDARD; PRT; 196 AA.  
P27657;  
01-AUG-1992 (Rel. 23, Created)  
01-AUG-1992 (Rel. 23, Last sequence update)  
28-FEB-2003 (Rel. 41, Last annotation update)  
Transcriptional regulatory protein uhpA.  
UHPA OR STM3790.  
Salmonella typhimurium.  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Salmonella.  
NCBI\_TaxID=602;  
[1]  
STRAIN=LT2;  
MEDLINE=92234930; PubMed=1569007;  
Island M.D.; Wei B.-Y.; Kadner R.J.;  
"Structure and function of the uhp genes for the sugar phosphate  
transport system in Escherichia coli and Salmonella typhimurium";  
J. Bacteriol. 174:2754-2762 (1992).  
[2]  
SEQUENCE FROM N.A.  
STRAIN=LT2 / SGSC412 / ATCC 700720;  
MEDLINE=21534948; PubMed=11677609;  
McClelland M., Sanderson K.B., Spieth J., Clifton S.W., Latreille P.,

RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,  
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
Ryan E., Sun H., Flores L., Miller W., Stoneking I., Nhan M.,  
Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2";  
RL Nature 413:852-856 (2001).  
CC -!- FUNCTION: APPEARS TO BE THE POSITIVE ACTIVATOR OF UHP  
TRANSCRIPTION.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -!- PTM: PHOSPHORYLATED BY UHP (PROBABLE).  
CC -!- SIMILARITY: Contains 1 response regulatory domain.  
CC -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL  
REGULATORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M89480; AAA27243.1; -  
EMBL; AE008876; AAL26448.1; -  
PIR; A41853; A41853.  
HSSP; P10957; IRNL.  
DR StyGene; SG10406; uhpA.

DR InterPro; IPR000792; HTH LuxR.  
DR InterPro; IPR001789; Response\_reg.  
Pfam; PF00196; GcrE; 1.  
Pfam; PF00072; response\_reg; 1.  
PRINTS; PR00038; HTHLUXR.  
ProDom; PD000307; HTH LuxR; 1.  
ProDom; PD000039; Response\_reg; 1.  
SMART; SM00421; HTH LUXR; 1.  
SMART; SM00448; REC\_1.  
PROSITE; PS00622; HTH LUXR FAMILY; 1.  
PROSITE; PS0110; RESPONSE REGULATORY; 1.  
Sensory transduction; Phosphorylation; Transcription regulation;  
DNA-binding; Activator; Sugar transport; Complete proteome.  
DOMAIN 1 116 RESPONSE REGULATORY.  
MOD RES 54 54 PHOSPHORYLATION (BY SIMILARITY).  
DNA\_BIND 155 174 H-T-H MOTIF (BY SIMILARITY).  
SEQUENCE 196 AA; 20790 MW; 4C80FF931C69078F CRC64;

Query Match 100.0%; Score 19; DB 1; Length 196;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 30 VAEF 33

RESULT 53  
AICD\_MOUSE STANDARD; PRT; 198 AA.  
ID AICD\_MOUSE  
AC Q9WE0;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Activation-induced cytidine deaminase (BC 3.5.4.5) (Cytidine  
aminohydrolase).  
GN AICDA OR AID.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99303612; PubMed=10373455;  
Muramatsu N., Sankaranand V.S., Anant S., Sugai M., Kinoshita K.,

RA Davidson N.O., Honjo T.;  
 RT "Specific expression of activation-induced cytidine deaminase (AID), a  
 RT novel member of the RNA-editing deaminase family in germinal center B  
 RT cells.";  
 RL J. Biol. Chem. 274:18470-18476(1999).  
 CC - FUNCTION: RNA-editing deaminase involved in somatic hypermutation,  
 CC gene conversion, and class-switch recombination. Required for  
 CC several crucial steps of B-cell terminal differentiation necessary  
 CC for efficient antibody responses.  
 CC - CATALYTIC ACTIVITY: Cytidine + H(2)O = uridine + NH(3).  
 CC - COFACTOR: Zinc (by similarity).  
 CC - SIMILARITY: Belongs to the cytidine and deoxycytidylate deaminases  
 CC family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF132979; AAD41793.1; --  
 CC MGD; MGI:1342279; Aicda.  
 CC GO; GO:0004126; F:cytidine deaminase activity; IDA.  
 CC InterPro; IPR007904; APOBEC C.  
 CC InterPro; IPR002125; dCMP/cyt\_deam.  
 CC Pfam; PF05240; APOBEC C; 1.  
 CC PROSITE; PS00903; Cyt\_DCMP\_DEAMINASES; 1.  
 CC mRNA processing; Hydroxase; Zinc.  
 CC METAL 56 56 ZINC (BY SIMILARITY).  
 CC METAL 87 87 ZINC (BY SIMILARITY).  
 CC METAL 90 90 ZINC (BY SIMILARITY).  
 CC SEQUENCE 198 aa; 24030 MW; 18A3BA10CA54BBE2 CRC64;  
 CC  
 CC Query Match 100.0%; Score 19; DB 1; Length 198;  
 CC Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 CC Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC Db 94 VAEF 97  
 CC  
 CC RESULT 54  
 CC HIS5 SULSO  
 CC ID HIS5 SULSO STANDARD; PRT; 199 AA.  
 CC AC Q3377;  
 CC DT 15-DEC-1998 (Rel. 37, Created)  
 CC DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Imidazole glycerol phosphate synthase subunit hsh (EC 2.4.2.-) (IGP  
 CC synthase glutamine amidotransferase subunit) (IGP synthase subunit  
 CC hsh) (ImGP synthase subunit hsh) (IGPS subunit hsh).  
 CC DE HSH OR SS00600 OR C08\_050.  
 CC GN Sulfolobus solfataricus.  
 CC OS Sulfolobus solfataricus.  
 CC OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 CC Sulfolobus.  
 CC OX NCBI\_TaxID=2287;  
 CC [1]  
 CC RN SEQUENCE FROM N.A.  
 CC RP STRAIN=ATCC 35092 / DSM 1617 / P2;  
 CC RX MEDLINE=97352708; PubMed=9209067;  
 CC RA Charlebois R.L., Senses C.W., Doolittle W.F., Brown J.R.;  
 RT "Evolutionary analysis of the hscGRdFDEH1 gene cluster from the  
 RT archaeon Sulfolobus solfataricus P2.";  
 RL J. Bacteriol. 179:4429-4432(1997).  
 CC [2]  
 CC RN SEQUENCE FROM N.A.  
 CC RP STRAIN=ATCC 35092 / DSM 1617 / P2;  
 CC RX MEDLINE=20165948; PubMed=10701121;  
 CC RA Charlebois R.L., Singh R.K., Chan-Weiher C.C.-Y., Allard G., Chow C.,  
 CC Confalonieri F., Curtis B., Brauso M., Duguet M., Faguy D.,

RA Gaasterland T., Garrett R.A., Gordon P., Jeffries A.C., Kozera C.,  
 RA Kusbaha N., Lafleur E., Medina N., Peng X., Penny S.L., She Q.,  
 RA St Jean A., van der Oost J., Young F., Zivanovic Y., Doolittle W.F.,  
 RA Ragan M.A., Senses C.W.;  
 RT "Gene content and organization of a 281-kbp contig from the genome of  
 RT the extremely thermophilic archaeon, Sulfolobus solfataricus P2.";  
 RL Genome 43:116-136(2000).  
 CC [3]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 CC MEDLINE=21332296; PubMed=11427726;  
 CC She Q., Singh R.K., Confalonieri F., Zivanovic Y., Curtis B.A.,  
 CC Aweez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Gordon P.M.K.,  
 CC De Moors A., Brauso G., Fletcher C., Kozera C.J., Medina N., Peng X.,  
 CC Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Tolstrup N.,  
 CC Thirumangalakudi V., Redder P., Schenk M.E., Theriault C.,  
 CC Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 CC Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.,  
 CC "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 CC - FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to  
 CC IGP, AICAR and glutamate. The hsh subunit provides the ammonia necessary to  
 CC amidotransferase activity that produces the ammonia necessary to  
 CC hshF for the synthesis of IGP and AICAR (By similarity).  
 CC - CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-  
 CC ylaminomethylideneamino)-1-(5-phosphoribosyl)imidazole-4-  
 CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-  
 CC aminimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.  
 CC - PATHWAY: Histidine biosynthesis; fifth step.  
 CC - SUBUNIT: Heterodimer of hsh and hshF (By similarity).  
 CC - SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC - SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U82227; AAB63025.1; --  
 CC EMBL; Y18930; CAB57700.1; --  
 CC EMBL; AE006689; AAK40911.1; --  
 CC PIR; H90206; H90206.  
 CC HAMAP; MF\_00278; -- 1.  
 CC InterPro; IPR000991; GATase\_1.  
 CC Pfam; PF00117; GATase; 1.  
 CC PROSITE; PS00442; GATASE\_TYPE I; FALSE NEG.  
 CC KW Histidine biosynthesis; Transferase; Glutamine amidotransferase;  
 CC Complete proteome.  
 CC ACT\_SITE 76 76 BY SIMILARITY.  
 CC ACT\_SITE 178 178 BY SIMILARITY.  
 CC ACT\_SITE 180 180 BY SIMILARITY.  
 CC SEQUENCE 199 AA; 22533 MW; 17DB1A86724CE095 CRC64;  
 CC  
 CC Query Match 100.0%; Score 19; DB 1; Length 199;  
 CC Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 CC Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 VAEF 4  
 CC DB 51 VAEF 54  
 CC  
 CC RESULT 55  
 CC RETB\_HORSE  
 CC ID RETB\_HORSE STANDARD; PRT; 201 AA.  
 CC AC Q28369;  
 CC DT 28-FEB-2003 (Rel. 41, Created)  
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Plasma retinol-binding protein precursor (PRBP) (RBP).

Y RBP4.  
 S Equus caballus (Horse).  
 C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 NCBI\_TaxID=9796;  
 (1) TAXID=9796;  
 SEQUENCE FROM N.A.  
 TISSUE=Endometrium;  
 MEDLINE=95226610; PubMed=7936053;  
 McDowell K.J., Adams M.H., Franklin K.M., Baker C.B.;  
 "Changes in equine endometrial retinol-binding protein RNA during the  
 estrous cycle and early pregnancy and with exogenous steroids.";  
 Biol. Reprod. 52:438-443(1995).  
 -!- FUNCTION: Delivers retinol from the liver stores to the peripheral  
 tissues. In plasma, the RBP-retinol complex interacts with  
 transthyretin, this prevents its loss by filtration through the  
 kidney glomeruli.  
 -!- SUBCELLULAR LOCATION: Secreted.  
 -!- SIMILARITY: Belongs to the lipocalin family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.ebi.ac.uk/submit/>)  
 or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk).  
 -----  
 EMBL; U21208; AAC48461.1; -;  
 PIR; I46257; I46257.  
 HSP; P27485; IAOB.  
 InterPro; IPR002345; Lipocalin.  
 InterPro; IPR000566; Lipocalin\_cycPBP.  
 Pfam; PF00061; Lipocalin; 1.  
 PRINTS; PR00179; LIPOCALIN.  
 PROSITE; PS00213; LIPOCALIN.  
 Plasma; Vitamin A; Retinol-binding; Transport; Liver; Signal;  
 Lipocalin.  
 SIGNAL 1 18 POTENTIAL.  
 CHAIN 19 201 PLASMA RETINOL-BINDING PROTEIN.  
 DISULFID 22 178 BY SIMILARITY.  
 DISULFID 88 192 BY SIMILARITY.  
 DISULFID 138 147 BY SIMILARITY.  
 I T SEQUENCE 201 AA; 23022 MW; 12CF80834E4262DC CRC64;  
 Query Match 100.0%; Score 19; DB 1; Length 201;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02; Indels 0; Gaps 0;  
 Matches 4; Conservative 0; Mismatches 0; Mismatches 0; Gaps 0;  
 Y 1 VAEF 4  
 b 60 VAEF 63  
 |||||  
 RESULT 56  
 ETE HUMAN  
 D RETE HUMAN STANDARD; PRT; 201 AA.  
 C P02753; O43478; O43479; Q8WNA3; Q9P178;  
 T 21-JUL-1986 (Rel. 01, Created)  
 T 28-FEB-2003 (Rel. 41, Last sequence update)  
 T 10-OCT-2003 (Rel. 42, Last annotation update)  
 N Plasma retinol-binding protein precursor (PRBP) (RBP) (PRO22222).  
 E RBP4.  
 S Homo sapiens (Human).  
 C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 (1) TAXID=9606;  
 SEQUENCE FROM N.A.  
 MEDLINE=84069802; PubMed=6316270;  
 Colantuoni V., Romano V., Bensi G., Santoro C., Costanzo F.,  
 Raugi G., Cortese R.;  
 "Cloning and sequencing of a full length cDNA coding for human

RT retinol-binding protein.";  
 RL Nucleic Acids Res. 11:7769-7776(1983).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,  
 RA Diatchenko L., Marzina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stalchenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Marra M.A.;  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RL [3]  
 RN SEQUENCE OF 1-189 FROM N.A.  
 RP MEDLINE=8605755; PubMed=2998779;  
 RX D'Onofrio C., Colantuoni V., Cortese R.;  
 RA "Structure and cell-specific expression of a cloned human retinol  
 RT binding protein gene: the 5'-flanking region contains hepatoma  
 RT specific transcriptional signals.";  
 RL EMBO J. 4:1981-1989(1985).  
 RN [4]  
 RP SEQUENCE OF 19-201, AND DISULFIDE BONDS.  
 RX MEDLINE=88019004; PubMed=2444024;  
 RA Risk L., Anundi H., Fohlman J., Peterson P.A.;  
 RT "The complete amino acid sequence of human serum retinol-binding  
 RT protein.";  
 RL Ups. J. Med. Sci. 92:115-146(1987).  
 RN [5]  
 RP SEQUENCE OF 19-201.  
 RX MEDLINE=81254137; PubMed=6942701;  
 RA Risk L., Anundi H., Boehme J., Eriksson U., Ronne H., Sege K.,  
 RA Peterson P.A.;  
 RT "Structural and functional studies of vitamin A-binding proteins.";  
 RL Ann. N.Y. Acad. Sci. 359:79-90(1981).  
 RN [6]  
 RP SEQUENCE OF 19-183.  
 RX MEDLINE=80004132; PubMed=573217;  
 RA Risk L., Anundi H., Peterson P.A.;  
 RT "The primary structure of the human retinol-binding protein.";  
 RL FEBS Lett. 104:55-58(1979).  
 RN [7]  
 RP SEQUENCE OF 18-201 FROM N.A.  
 RC TISSUE=Fetal liver;  
 RA Zhang C., Yu Y., Zhang S., Wei H., Zhang Y., Zhou G., Bi J., Liu M.,  
 RA He F.;  
 RT "Functional prediction of the coding sequences of 79 new genes deduced  
 RT by analysis of cDNA clones from human fetal liver.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE=95395382; PubMed=7666002;  
 RA Jacot S., Rose K., Hughes G.J., Saurat J.-H., Siegenthaler G.;  
 RT "Characterization of two post-translationally processed forms of  
 RT human serum retinol-binding protein: altered ratios in chronic renal  
 RT failure.";  
 RL J. Lipid Res. 36:1247-1253(1995).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=91017498; PubMed=2217163;

RA Cowan S.W., Newcomer M.E., Jones T.A.;  
RT "Crystallographic refinement of human serum retinol binding protein  
RL at 2-A resolution."  
RN Proteins 8:44-61(1990).  
[10]  
RP X-RAY CRYSTALLOGRAPHY.  
RX MEDLINE=92322903; PubMed=1623143;  
RA Monaco H.L., Zanotti G.;  
RT "Three-dimensional structure and active site of three hydrophobic  
RT molecule-binding proteins with significant amino acid sequence  
RT similarity."  
RL Biopolymers 32:457-465(1992).  
[11]  
RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS) OF COMPLEX WITH TTR.  
RX MEDLINE=99162254; PubMed=10052934;  
RA Naylor H.M., Newcomer M.E.;  
RT "The structure of human retinol-binding protein (RBP) with its  
RT carrier protein transthyretin reveals an interaction with the carboxy  
RT terminus of RBP."  
RL Biochemistry 38:2647-2653(1999).  
[12]  
RP VARIANTS RBP DEFICIENCY ASN-59 AND ASP-93.  
RX MEDLINE=99103495; PubMed=9888420;  
RA Seeliger M.W., Biesalski H.K., Wessinger B., Gollnick H., Gielen S.,  
RA Frank J., Beck S., Zrenner E.;  
RT "Phenotype in retinol deficiency due to a hereditary defect in retinol  
RT binding protein synthesis."  
RL Invest. Ophthalmol. Vis. Sci. 40:3-11(1999).  
[13]  
RP CHARACTERIZATION OF VARIANTS ASN-59 AND ASP-93.  
RX MEDLINE=99247593; PubMed=10232633;  
RA Biesalski H.K., Frank J., Beck S.C., Heinrich F., Illek B., Reifen R.,  
RA Gollnick H., Seeliger M.W., Wessinger B., Zrenner E.;  
RT "Biochemical but not clinical vitamin A deficiency results from  
RT mutations in the gene for retinol binding protein."  
RL Am. J. Clin. Nutr. 69:931-936(1999).  
CC -!- FUNCTION: Delivers retinol from the liver stores to the peripheral  
CC tissues. In plasma, the RBP-retinol complex interacts with  
CC transthyretin, this prevents its loss by filtration through the  
CC kidney glomeruli.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- MASS SPECTROMETRY: MW=21063.46; MW\_PRR=1.88; METHOD=Electrospray;  
CC RANGE=17-199.  
CC -!- DISEASE: Defects in RBP4 are a cause of retinol-binding protein  
CC deficiency [MIM:180250]. This condition causes night vision  
CC problems. It produces a typical "fundus xerophthalmicus,"  
CC featuring a progressed atrophy of the retinal pigment epithelium.  
CC -!- DISEASE: A deficiency of vitamin A blocks secretion of the binding  
CC protein posttranslationally and results in defective delivery and  
CC supply of vitamin to the epidermal cells (a condition associated  
CC with a dermatosis).  
CC -!- SIMILARITY: Belongs to the lipocalin family.  
CC -!- DATABASE: NAME=Mutations of the RBP4 gene.  
CC NOTE=Retina International's Scientific Newsletter:  
CC WWW="http://www.retina-international.com/sci-news/ribsp4mut.htm".  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; X00129; CAA24959.1; --  
CC EMBL; BC020633; AAH20633.1; --  
CC EMBL; X02775; CAA26553.1; --  
CC EMBL; X02824; CAB46489.1; --  
CC EMBL; AF119868; AAF69622.1; ALT\_INIT.  
CC EMBL; AF025334; AAC02945.1; --  
CC EMBL; AF025335; AAC02946.1; --  
CC PIR; A93494; VAHU.  
CC DB; 1RBP; 15-JUL-91.

DR PDB; 1BRP; 31-JAN-94.  
DR PDB; 1BRQ; 31-JAN-94.  
DR PDB; 1QAB; 09-APR-99.  
DR SWISS-2DPAGE; P02753; HUMAN.  
DR Siena-2DPAGE; P02753; --  
DR Genew; HGNC:9922; RBP4.  
DR MIW; 180250; --  
DR GO; GO:0005615; C:extracellular space; TAS.  
DR GO; GO:0005501; F:retinoid binding; TAS.  
DR InterPro; IPR002345; Lipocalin.  
DR InterPro; IPR000566; Lipocln\_cytFABP.  
DR Pfam; PF00061; Lipocalin; 1.  
DR PRINTS; PS00179; LIPOCALIN.  
DR PROSITE; PS00213; LIPOCALIN; 1.  
KW Plasma; Vitamin A; Retinol-binding; Transport; Liver; Signal;  
KW Lipocalin; Disease mutation; Vision; 3D-structure.  
FT SIGNAL 1 18  
FT CHAIN 19 201 PLASMA RETINOL-BINDING PROTEIN.  
FT DISULFID 22 178  
FT DISULFID 88 192  
FT DISULFID 138 147  
FT VARIANT 59 59  
FT VARIANT 93 93  
FT VARIANT 8 8  
FT CONFLICT 13 17  
FT CONFLICT 23 23  
FT STRAND 24 26  
FT HELIX 31 32  
FT TURN 35 38  
FT STRAND 40 48  
FT STRAND 60 66  
FT TURN 67 69  
FT STRAND 70 80  
FT TURN 82 83  
FT STRAND 86 96  
Query Match 100.0%; Score 19; DB 1; Length 201;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
DB 60 VAEF 63  
RESULT 57  
RETB\_PIG  
ID RETB\_PIG STANDARD; PRT; 201 AA.  
AC P27485;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Plasma retinol-binding protein precursor (PRBP) (RBP).  
GN RBP4.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92131004; PubMed=1723146;  
RA Trout W.E., McDonnell J.J., Kramer K.K., Raumbach G.A., Roberts R.M.;  
RT "The retinol-binding protein of the expanding pig blastocyst:  
RT molecular cloning and expression in trophectoderm and embryonic  
RT disc."  
RL Mol. Endocrinol. 5:1533-1540(1991).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS), AND REVISIONS TO 134 AND 185.  
RX MEDLINE=98437649; PubMed=9757135;  
RA Zanotti G., Panzavolta M., Marcato A., Malpeli G., Polli C., Berni R.;  
RT "Structure of pig plasma retinol-binding protein at 1.65-A

```

I resolution."
L Acta Crystallogr. D 54:1049-1052(1998).
C -I- FUNCTION: Delivers retinol from the liver stores to the peripheral
C tissues. In plasma, the RBP-retinol complex interacts with
C tranthyretin, this prevents its loss by filtration through the
C kidney glomeruli.
C -I- SUBCELLULAR LOCATION: Secreted.
C -I- SIMILARITY: Belongs to the lipocalin family.
C -----
C This SWISS-PROT entry is copyrighted. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
C EMBL; M68860; AAA31113.1; -.
C PIR; A39486; A39486.
C PDB; 1AOB; 28-JAN-98.
C InterPro; IPR002345; Lipocalin.
C Pfam; PF00061; Lipocalin; 1.
C PRINTS; PR00179; LIPOCALIN.
C PROSITE; PS00213; LIPOCALIN; 1.
C Plasma; Vitamin A; Retinol-binding; Transport; Liver; Signal;
C Lipocalin; 3D-structure.
C SIGNAL 1 18
C CHAIN 19 201 POTENTIAL.
C DISULFID 22 178 PLASMA RETINOL-BINDING PROTEIN.
C DISULFID 88 192
C DISULFID 138 147
C CONFLICT 134 134 V -> A (IN REF. 1).
C CONFLICT 185 185 I -> L (IN REF. 1).
C STRAND 23 23
C HELIX 24 26
C TURN 31 32
C HELIX 35 38
C STRAND 40 48
C STRAND 55 65
C TURN 67 68
C STRAND 71 80
C TURN 82 83
C STRAND 86 97
C TURN 101 102
C STRAND 103 110
C TURN 113 114
C STRAND 118 127
C STRAND 132 141
C TURN 143 144
C STRAND 146 156
C TURN 159 160
C HELIX 164 176
C TURN 177 178
C TURN 180 181
C STRAND 184 185
C TURN 191 192
C SEQUENCE 201 AA; 23067 MW; A20E39D3C9471DC8 CRC64;
C -----
C Query Match 100.0%; Score 19; DB 1; Length 201;
C Best Local Similarity 100.0%; Pred. No. 3.5e+02;
C Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
C -----
C Y 1 VAEF 4
C |
C |
C |
C 60 VAEF 63
C -----
C RESULT 58
C ETE RABIT STANDARD; PRT; 201 AA.
C D RETB RABIT
C P06912;
C 01-JAN-1988 (Rel. 06, Created)

```

```

DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plasma retinol-binding protein precursor (PRBP) (RBP).
GN RBP4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=93011736; PubMed=1339354;
RX Lee S.Y., Uebels J.L., Soprano D.R.;
RA "The lacrimal gland synthesizes retinol-binding protein.";
RL Exp. Eye Res. 55:163-171(1992).
RN [2]
RN SEQUENCE OF 19-201.
RP MEDLINE=85207643; PubMed=3838985;
RX Sundelin J., Laurent B.C., Anundi H., Traegaardh L., Larhammar D.,
RA Bjoerck L., Eriksson U., Aakerstroem B., Jones A., Newcomer M.,
RA Peterson P.A., Rask L.;
RT "Amino acid sequence homologies between rabbit, rat, and human serum
RT retinol-binding proteins.";
RL J. Biol. Chem. 260:6472-6480(1985).
CC -I- FUNCTION: Delivers retinol from the liver stores to the peripheral
CC tissues. In plasma, the RBP-retinol complex interacts with
CC tranthyretin, this prevents its loss by filtration through the
CC kidney glomeruli.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: Belongs to the lipocalin family.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S45958; AAB23592.1; -.
CC PIR; A49178; VAB.
CC HSP; P18902; 1FEN.
CC InterPro; IPR002345; Lipocalin.
CC InterPro; IPR000566; Lipocln_cytpabp.
CC Pfam; PF00061; Lipocalin; 1.
CC PRINTS; PR00179; LIPOCALIN.
CC PROSITE; PS00213; LIPOCALIN; 1.
CC Plasma; Vitamin A; Retinol-binding; Transport; Liver; Lipocalin;
CC Signal.
CC SIGNAL 1 18
CC CHAIN 19 201 PLASMA RETINOL-BINDING PROTEIN.
CC DISULFID 22 178 BY SIMILARITY.
CC DISULFID 88 192 BY SIMILARITY.
CC DISULFID 138 147 BY SIMILARITY.
CC SEQUENCE 201 AA; 23102 MW; 4153FCF050184136 CRC64;
C -----
C Query Match 100.0%; Score 19; DB 1; Length 201;
C Best Local Similarity 100.0%; Pred. No. 3.5e+02;
C Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
C -----
C Y 1 VAEF 4
C |
C |
C |
C 60 VAEF 63
C -----
C RESULT 59
C YIGZ ECOLI STANDARD; PRT; 204 AA.
C ID YIGZ ECOLI
C AC P27862;
C 01-AUG-1992 (Rel. 23, Created)
C 15-OCT-2001 (Rel. 40, Last sequence update)
C 18-OCT-2001 (Rel. 40, Last annotation update)
C DE Hypothetical protein YIGZ.
C YIGZ OR B3848.

```

```

JS Escherichia coli.
XC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
XC Enterobacteriaceae; Escherichia.
XX NCBI_TaxID=562;
XN [1]
XP SEQUENCE FROM N.A.
XC STRAIN=K12 / MG1655;
XA MEDLINE=92358234; PubMed=1379743;
LA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
XT "Analysis of the Escherichia coli genome: DNA sequence of the region
TL from 84.5 to 86.5 minutes.";
XN Science 257:771-778(1992).
XN [2]
XP SEQUENCE FROM N.A.
XC STRAIN=K12;
XA MEDLINE=91057145; PubMed=2243799;
LA Nakahigashi K., Inokuchi H.;
XT "Nucleotide sequence between the fadB gene and the rna operon from
XN Escherichia coli.";
XN Nucleic Acids Res. 18:6439-6439(1990).
XN [3]
XP SIMILARITY.
XA MEDLINE=94147996; PubMed=8313894;
LA Koonin E.V., Bork P., Sander C.;
XT "Yeast chromosome III: new gene functions.";
XN EMBO J. 13:493-503(1994).
XC -1- SIMILARITY: BELONGS TO THE UPF0029 FAMILY. STRONG, TO H.INFLUENZAE
XN H10722.
XC -----
XC This SWISS-PROT entry is copyright. It is produced through a collaboration
XC between the Swiss Institute of Bioinformatics and the EMBL outstation -
XC the European Bioinformatics Institute. There are no restrictions on its
XC use by non-profit institutions as long as its content is in no way
XC modified and this statement is not removed. Usage by and for commercial
XC entities requires a license agreement (see http://www.isb-sib.ch/announce/
XC or send an email to license@isb-sib.ch).
XC -----
XN EMBL; M87049; AAA67645.1; ALT INIT.
XN EMBL; AE000460; AAC76851.1; ALT INIT.
XN EMBL; X54687; CAA38501.1; ALT_INIT.
XN EcoGene; EGI1484; Y192.
XN InterPro; IPR009022; EFG_III_V.
XN InterPro; IPR001498; UPF0029.
XN Pfam; PF01205; UPF0029; 1.
XN TIGRFAMs; TIGR00257; TIGR00257; 1.
XN PROSITE; PS00910; UPF0029; 1.
XN Hypothetical protein; Complete proteome.
XN CONFLICT 125 125 L -> V (IN REF. 2).
XN SEQUENCE 204 AA; 21757 MW; 97E48163FE899DBE CRC64;
XQ QUERY MATCH
XQ Best Local Similarity 100.0%; Score 19; DB 1; Length 204;
XQ Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XY 1 VAEF 4
XB 182 VAEF 185
XZ RESULT 60
XD21 HUMAN
XD MD21 HUMAN STANDARD; PRT; 205 AA.
XC Q13257;
XJ 16-OCT-2001 (Rel. 40, Created)
XJ 16-OCT-2001 (Rel. 40, Last sequence update)
XJ 10-OCT-2003 (Rel. 42, Last annotation update)
XJ Mitotic spindle assembly checkpoint protein MAD2A (MAD2-like 1)
XD (HsMAD2).
XN MAD2L1 OR MAD2.
XN Homo sapiens (Human).
XC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
XC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
XN NCBI_TaxID=9606;

```

```

RN RP SEQUENCE FROM N.A.
RX MEDLINE=96421709; PubMed=8824189;
RA Li Y., Benezra R.;
RT "Identification of a human mitotic checkpoint gene: hSMAD2.";
RL Science 274:246-248(1996).
RN [2]
XP SEQUENCE FROM N.A.
RA Jin D.-Y., Jeang K.-T.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
XP SEQUENCE FROM N.A.
RA Klebert S., Barnikol-Watanabe S., Kratzin H.D., Hilschmann N.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
XP SEQUENCE FROM N.A.
RA Nobori T.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
XP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow, and Muscle;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedlin T.B., Ioshizuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellan N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McElwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Viallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
XP INTERACTION WITH CDC20.
RX MEDLINE=98301442; PubMed=9637688;
RA Fang G., Yu H., Kirschner M.W.;
RT "The checkpoint protein MAD2 and the mitotic regulator CDC20 form a
RT ternary complex with the anaphase-promoting complex to control
RT anaphase initiation.";
RL Genes Dev. 12:1871-1883(1998).
RN [7]
XP INTERACTION WITH ADAM17.
RX MEDLINE=99458684; PubMed=10527948;
RA Nelson K.K., Schlondorff J., Blobel C.P.;
RT "Evidence for an interaction of the metalloprotease-disintegrin tumour
RT necrosis factor alpha convertase (TACE) with mitotic arrest deficient
RT 2 (MAD2), and of the metalloprotease-disintegrin MDC9 with a novel
RT MAD2-related protein, MAD2-beta.";
RL Biochem. J. 343:673-680(1999).
RN [8]
XP STRUCTURE BY NMR OF 11-195.
RX MEDLINE=20165182; PubMed=10700282;
RA Luo X., Fang G., Coldiron M., Lin Y., Yu H., Kirschner M.W.,
RA Wagner G.;
RT "Structure of the Mad2 spindle assembly checkpoint protein and its
RT interaction with Cdc20.";
RL Nat. Struct. Biol. 7:224-229(2000).
XC -1- FUNCTION: REQUIRED FOR THE EXECUTION OF THE MITOTIC CHECKPOINT
XC WHICH MONITORS THE PROCESS OF KINETOCORE-SPINDLE ATTACHMENT AND
XC DELAYS THE ONSET OF ANAPHASE WHEN THIS PROCESS IS NOT COMPLETE. IT
XC INHIBITS THE ACTIVITY OF THE ANAPHASE PROMOTING COMPLEX BY
XC SEQUESTERING CDC20 UNTIL ALL CHROMOSOMES ARE ALIGNED AT THE
XC METAPHASE PLATE.

```

-----  
-!- SUBUNIT: INTERACTS WITH CDC20 AND WITH ADAM17 (TACE).  
-!- SUBCELLULAR LOCATION: Nuclear.  
-!- SIMILARITY: BELONGS TO THE MAD2 FAMILY.  
-!- SIMILARITY: CONTAINS 1 HORMA DOMAIN.  
-!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;  
WWW="http://www.infobiogen.fr/services/Chromcancer/Genes/MAD2L1ID304.html".  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U65410; AAC50781.1; -.  
EMBL; U31278; AAC52060.1; -.  
EMBL; AJ000186; CAA03943.1; -.  
EMBL; AB056160; BAB63410.1; -.  
EMBL; BC000356; BAH00356.1; -.  
EMBL; BC005945; AAH05945.1; -.  
PIR; G01942; G01942.  
PDB; 1DUJ; 08-MAR-00.  
Genew; HGNC:6763; MAD2L1.  
GK; Q13257; -.  
MIM; 601467; -.  
GO; GO:0005699; C:kinetochore; TAS.  
GO; GO:0007067; P:mitosis; TAS.  
GO; GO:0007093; P:mitotic checkpoint; TAS.  
InterPro; IPR003511; DNABind\_HORMA.  
Pfam; PF02301; HORMA; 1.  
PROSITE; PS50815; HORMA; 1.  
Cell cycle; Mitosis; Nuclear protein; 3D-structure.  
DOMAIN 14 197  
STRAND 11 12  
STRAND 15 15  
HELIX 17 34  
TURN 35 35  
TURN 41 42  
TURN 44 48  
TURN 49 50  
TURN 51 55  
TURN 59 73  
STRAND 83 90  
TURN 91 94  
STRAND 95 103  
STRAND 106 106  
STRAND 112 113  
HELIX 122 141  
STRAND 149 158  
TURN 165 166  
TURN 173 175  
TURN 179 181  
TURN 185 193  
SEQUENCE 205 AA; B8DCBF0043836764 CRC64;  
Query Match 100.0%; Score 19; DB 1; Length 205;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 VAEF 4  
b 20 VAEF 23  
-----  
RESULT 61  
D21\_MOUSE  
D MAD21\_MOUSE STANDARD; PRT; 205 AA.  
C Q521B5; Q9J153;  
T 16-OCT-2001 (Rel. 40, Created)  
T 16-OCT-2001 (Rel. 40, Last sequence update)  
T 10-OCT-2003 (Rel. 42, Last annotation update)  
E Mitotic spindle assembly checkpoint protein MAD2A (MAD2-like 1).  
-----

GN MAD2L1 OR MAD2A.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryo;  
RA Jin D.-Y.; Jeang K.-T.;  
RT Identification of a novel component of the spindle assembly  
RT checkpoint in mammalian cells.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129;  
RX MEDLINE=20349739; PubMed=10892650;  
RA Dobles M., Liberal V., Scott M.L., Benzer R., Sorger P.K.;  
RT "Chromosome missegregation and apoptosis in mice lacking the mitotic  
RT checkpoint protein Mad2.";  
RL Cell 101:635-645(2000).  
CC -!- FUNCTION: REQUIRED FOR THE EXECUTION OF THE MITOTIC CHECKPOINT  
CC DELAYS THE ONSET OF ANAPHASE WHEN THIS PROCESS IS NOT COMPLETE. IT  
CC INHIBITS THE ACTIVITY OF THE ANAPHASE PROMOTING COMPLEX BY  
CC SEQUESTERING CDC20 UNTIL ALL CHROMOSOMES ARE ALIGNED AT THE  
CC METAPHASE PLATE (BY SIMILARITY).  
CC -!- SUBUNIT: INTERACTS WITH CDC20.  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE MAD2 FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 HORMA DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U83902; AAD09238.1; -.  
DR EMBL; AF261919; AAF69525.1; -.  
DR HSSP; Q13257; 1DUJ.  
DR MGD; MGI:1860374; Mad2l1.  
DR GO; GO:000776; C:kinetochore; IDA.  
DR GO; GO:0005815; C:spindle pole body; IDA.  
DR GO; GO:0000070; P:mitotic chromosome segregation; IMP.  
DR GO; GO:0007094; P:mitotic spindle checkpoint; IMP.  
DR InterPro; IPR003511; DNABind\_HORMA.  
DR Pfam; PF02301; HORMA; 1.  
DR PROSITE; PS50815; HORMA; 1.  
KW Cell cycle; Mitosis; Nuclear protein.  
FT DOMAIN 14 197  
FT CONFLICT 157 157 T -> A (IN REF. 1).  
FT CONFLICT 178 178 C -> S (IN REF. 1).  
FT CONFLICT 201 201 T -> I (IN REF. 1).  
SQ SEQUENCE 205 AA; 23598 MW; A9F3F28BC4C9738E CRC64;  
Query Match 100.0%; Score 19; DB 1; Length 205;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
Db 20 VAEF 23  
-----  
RESULT 62  
YACL\_LEGPN  
ID YACL\_LEGPN STANDARD; PRT; 208 AA.  
AC P37033;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)



R Pfam; PF00760; Cucumo\_coat; 1.  
R PRINTS; PF00222; CUCUMOCOAT.  
R ProDom; PD001284; Cucumovirus\_coat; 1.  
W Coat protein.  
Q SEQUENCE 218 AA; 24185 MW; 130E82D17BD75224 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 218;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
b 96 VAEF 99

## RESULT 65

D TPI\_NPVOP STANDARD; PRT; 220 AA.  
C 01-NOV-1997 (Rel. 35, Created)  
T 01-NOV-1997 (Rel. 35, Last sequence update)  
I 10-OCT-2003 (Rel. 42, Last annotation update)  
E Putative protein-tyrosine phosphatase 1 (EC 3.1.3.48).  
N TPI-1.  
S Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMNPV).  
C Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
C Nucleopolyhedrovirus.  
X NCBI\_TaxID=164623;  
N [1]  
P SEQUENCE FROM N.A.  
F MEDLINE=97271300; PubMed=9126251;  
X Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,  
A Rohrmann G.F.;  
T "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear  
polyhedrosis virus genome.";  
L Virology 229:381-399(1997).  
C -!- FUNCTION: COULD BE INACTIVE AS THE ACTIVE SITE CYS-100 IS  
MODIFIED TO TRYPTOPHAN.  
C -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein  
tyrosine + phosphate.  
C -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
TYROSINE PHOSPHATASE FAMILY. CDC14 SUBFAMILY.

-----  
C This SWISS-PROT entry is copyright. It is produced through a collaboration  
C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
C the European Bioinformatics Institute. There are no restrictions on its  
C use by non-profit institutions as long as its content is in no way  
C modified and this statement is not removed. Usage by and for commercial  
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
R EMBL; U75930; AAC59009.1; -.  
R InterPro; IPR000387; TYR\_phosphatase.  
R PROSITE; PS00383; TYR\_PHOSPHATASE\_1; FALSE\_NEG.  
R PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
W Hydrolase.  
T SITE 162 162 EQUIVALENT OF ACTIVE SITE CYS.  
Q SEQUENCE 220 AA; 25234 MW; D3FC093F1953D425 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 220;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
b 139 VAEF 142

## RESULT 66

CRL\_ARCFU STANDARD; PRT; 223 AA.  
D ECR1\_ARCFU  
C O29759;  
T 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Probable exosome complex RNA-binding protein 1.  
GN AF0492.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., Hickey E.K., Peterson J.D.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Kyripides N.C.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Sutton G.G., Gill S.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Adams M.D., Loftus B.,  
RA Kirkness E.F., Dougherty B.A., McKenney K., Sutton G.G., Loftus B.,  
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,  
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370(1997).  
CC -!- FUNCTION: Probably involved in degradation of a variety of RNA  
species; could act a RNA-binding component of the exosome  
(Potential).  
CC -!- SUBUNIT: Component of the archaeal exosome multienzyme  
ribonuclease complex (Potential).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -!- SIMILARITY: Contains 1 KH domain.  
CC -!- SIMILARITY: Contains 1 S1 motif domain.  
-----  
C This SWISS-PROT entry is copyright. It is produced through a collaboration  
C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
C the European Bioinformatics Institute. There are no restrictions on its  
C use by non-profit institutions as long as its content is in no way  
C modified and this statement is not removed. Usage by and for commercial  
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
R EMBL; A5001070; AAB90745.1; -.  
R TIGR; AF0492; -.  
R HAMAP; MF\_00623; -; 1.  
R InterPro; IPR004087; KH dom.  
R InterPro; IPR004088; KH\_type\_1.  
R InterPro; IPR008994; Nucleic\_acid\_OB.  
R InterPro; IPR003029; S1.  
R Pfam; PF00013; KH; 1.  
R Pfam; PF00575; S1; 1.  
R SMART; SM00322; KH; 1.  
R SMART; SM00316; S1; 1.  
R PROSITE; PS50084; KH\_TYPE\_1; 1.  
R PROSITE; PS50126; S1; 1.  
R Exosome; RNA-binding; Complete proteome.  
FT DOMAIN 58 127 KH.  
FT DOMAIN 135 193 S1 MOTIF.  
SQ SEQUENCE 223 AA; 24725 MW; A11B7F1A26972833 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 223;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAEF 4  
Db 209 VAEF 212

## RESULT 67

RAN\_TETPY

```

ID  RAN_TETPY  STANDARD;  PRT;  223 AA.
AC  P41314;
DT  01-NOV-1995 (Rel. 32, Created)
DT  01-NOV-1995 (Rel. 32, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  GTP-binding nuclear protein RAN/TC4.
OS  Tetrahymena pyriformis.
OC  Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC  Tetrahymenina; Tetrahymena.
OX  NCBI_TaxID=5908;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=W;
RX  MEDLINE=94299154; PubMed=8026746;
RA  Nagata K.-I., Takemasa T., Alam S., Hattori T., Watanabe Y.,
RA  Nozawa Y.;
RT  "Cloning of cDNAs encoding a cell-cycle-regulatory GTP-binding
RT  low-M(r) (GBLM) protein, Ran/TC4, from micronucleated Tetrahymena
RT  thermophila and amiconucleated Tetrahymena pyriformis.";
RL  Gene 144:123-125 (1994).
CC  -!- FUNCTION: GTP-binding protein involved in nucleocytoplasmic
CC  transport. Required for the import of protein into the nucleus and
CC  also for RNA export. Involved in chromatin condensation and
CC  control of cell cycle (By similarity).
CC  -!- SUBUNIT: Monomer.
CC  -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC  -!- SIMILARITY: Belongs to the small GTPase superfamily. Ran family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; D21825; BAA04849.1; -.
DR  HSSP; P17080; 1A2K.
DR  InterPro; IPR002041; RAN.
DR  InterPro; IPR001806; Ras trnsfrmg.
DR  InterPro; IPR005225; Small_GTP.
DR  Pfam; PF00071; ras; 1.
DR  PRINTS; PRO0627; GTPRANTC4.
DR  PRINTS; PRO0449; RASTRNSFRMNG.
DR  SMART; SM00176; RAN; 1.
DR  TIGRFAMs; TIGR00231; small_GTP; 1.
DR  PROSITE; PS01115; RAN; 1.
KW  GTP-binding; Nuclear protein; protein transport.
FT  NP_BIND 18 25 GTP (BY SIMILARITY).
FT  NP_BIND 66 70 GTP (BY SIMILARITY).
FT  NP_BIND 123 126 GTP (BY SIMILARITY).
FT  DOMAIN 128 143 IBB (BY SIMILARITY).
SQ  SEQUENCE 223 AA; 25422 MW; 5474DADB5ABF0977 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.9e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0;

QY 1 VAEF 4
DB 9 VAEF 12

RESULT 68
VATD THETH STANDARD; PRT; 223 AA.
AC O87880;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE V-type ATP synthase subunit D (EC 3.6.3.14) (V-type ATPase subunit D).
GN ATPD OR VATD.
OS Thermus thermophilus.

```

```

OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HB8 / ATCC 27634;
RX MEDLINE=20250964; PubMed=10788522;
RA Yokoyama K., Okuma S., Taguchi H., Yasunaga T., Wakabayashi T.,
RA Yoshida M.;
RA "V-type H+-ATPase/synthase from a thermophilic eubacterium, Thermus
RT thermophilus. Subunit structure and operon.";
RL J. Biol. Chem. 275:13955-13961 (2000).
CC -!- FUNCTION: Produces ATP from ADP in the presence of a proton
CC gradient across the membrane.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SIMILARITY: Belongs to the V-ATPase D subunit family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63799; BAA33198.2; -.
DR HAVAP; MF 00271; -.
DR InterPro; IPR002699; ATPsynt_Dsub.
DR Pfam; PF01813; ATP_synt_D; 1.
DR ProDom; PD004122; ATPsynt_Dsub; 1.
DR TIGRFAMs; TIGR00309; V_ATPase_subD; 1.
KW Hydrolase; ATP synthase; Hydrogen ion transport.
SQ SEQUENCE 223 AA; 24677 MW; B666AE00DF0B08B CRC64;

Query Match 100.0%; Score 19; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.9e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0;

QY 1 VAEF 4
DB 36 VAEF 39

RESULT 69
RAN_TETTH STANDARD; PRT; 225 AA.
AC P41315;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE GTP-binding nuclear protein RAN/TC4.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=B1868;
RX MEDLINE=94299154; PubMed=8026746;
RA Nagata K.-I., Takemasa T., Alam S., Hattori T., Watanabe Y.,
RA Nozawa Y.;
RT "Cloning of cDNAs encoding a cell-cycle-regulatory GTP-binding
RT low-M(r) (GBLM) protein, Ran/TC4, from micronucleated Tetrahymena
RT thermophila and amiconucleated Tetrahymena pyriformis.";
RL Gene 144:123-125 (1994).
CC -!- FUNCTION: GTP-binding protein involved in nucleocytoplasmic
CC transport. Required for the import of protein into the nucleus and
CC also for RNA export. Involved in chromatin condensation and
CC control of cell cycle (By similarity).
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Belongs to the small GTPase superfamily. Ran family.
CC -----

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
-----  
EMBL; AE016792; AAN70379.1; -.  
TIGR; PP4910; -.  
HAMAP; MF 00244; -. 1.  
InterPro; IPR004820; Cysdilyltransf.  
Pfam; PF01457; CTP\_transf\_2; 1.  
Pyridine nucleotide biosynthesis; Transferase; Nucleotidyltransferase;  
NAB; Complete proteome.  
SEQUENCE 230 AA; 25291 MW; 4CF04E8F70B48941 CRC64;  
SQ

```

Query Match      100.0%; Score 19; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1 VAEF 4
3B 42 VAEF 45

RESULT 72
RE3_PORU
ID _RR3_PORPU STANDARD; PRT; 230 AA.
AC PS1308;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chloroplast 30S ribosomal protein S3.
GN RP83.
OS Porphyra purpurea.
OG Chloroplast.
CC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
CX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Avonport.
RA Reith M.E., Munnholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast genome."
RL Plant Mol. Biol. Rep. 13:333-335 (1995).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC
CC EVBL; U38804; AAC08194.1;
CC PIR; S73229; S73229.
CC HAMAP; MF 01309; -.
CC InterPro; IPR004087; KH_dom.
CC InterPro; IPR009019; KH_prok.
CC InterPro; IPR004044; KH_TYPE_2.
CC InterPro; IPR001351; Ribosomal_S3_C.
CC InterPro; IPR008282; Ribosomal_S3_N.
CC Pfam; PF00013; KH; 1.
CC Pfam; PF00189; Ribosomal_S3_C; 1.
CC Pfam; PF00417; Ribosomal_S3_N; 1.
CC SMART; SM00322; KH; 1.
CC TIGRFAMs; TIGR01009; rpsC_bact; 1.
CC PROSITE; PS00823; KH_TYPE_2; 1.
CC PROSITE; PS00548; RIBOSOMAL_S3; 1.
CC Ribosomal protein; RNA-binding; rRNA-binding; Chloroplast.
FT DOMAIN 39..109 KH TYPE-2.
SQ SEQUENCE 230 AA; 25631 MW; 5499B93704D7F0ED CRC64;

Query Match      100.0%; Score 19; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 118 VAEF 121

RESULT 73
UBIE_COREF

```

```

ID UBIE_COREF STANDARD; PRT; 230 AA.
AC Q8PSE3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Menaquinone biosynthesis methyltransferase ubiE (EC 2.1.1.-).
GN UBIE OR CR0481.
OS Corynebacterium efficiens.
CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
CC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
CX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RX MEDLINE=22723752; PubMed=12840036;
RA Nishio Y., Nakamura Y., Kawarabayashi Y., Usuda Y., Kimura E.,
RA Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,
RA Gojobori T.;
RA "Comparative complete genome sequence analysis of the amino acid
RT replacements responsible for the thermostability of Corynebacterium
RT efficiens."
RL Genome Res. 13:1572-1579 (2003).
CC -!- FUNCTION: Methyltransferase required for the conversion of
CC dimethylmenaquinone (DMK2) to menaquinone (MK2) (By similarity).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + demethylmenaquinol =
CC S-adenosyl-L-homocysteine + menaquinol.
CC -!- PATHWAY: Menaquinone biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the ubiE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC
CC EMBL; AP005215; BAC17291.1; ALT_INIT.
CC HAMAP; MF 01813; -.
CC InterPro; IPR001601; Methyltransf.
CC InterPro; IPR000051; SAM_bind.
CC InterPro; IPR004033; UbiE/COQ5_Metrf.
CC Pfam; PF01209; UbiE_methyltran; 1.
CC PROSITE; PS01183; UBI1; FALSE NEG.
CC PROSITE; PS01184; UBI2; FALSE NEG.
CC Menaquione biosynthesis; Transferase; Methyltransferase;
CC Complete proteome.
CC SEQUENCE 230 AA; 25366 MW; E10DCB602A1CD886 CRC64;

Query Match      100.0%; Score 19; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 145 VAEF 148

RESULT 74
UBIE_COREG
ID UBIE_COREG STANDARD; PRT; 230 AA.
AC Q8NT39;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Menaquione biosynthesis methyltransferase ubiE (EC 2.1.1.-).
GN UBIE OR CGL0471.
OS Corynebacterium glutamicum (Brevibacterium flavum).
CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
CC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
CX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.

```

STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
 A Makagawa S.;  
 "Complete genomic sequence of *Corynebacterium glutamicum* ATCC 13032.";  
 Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 -!- FUNCTION: Methyltransferase required for the conversion of  
 dimethylmenaquinone (DMK2) to menaquinone (MK2) (By similarity).  
 -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + demethylmenaquinol =  
 S-adenosyl-L-homocysteine + menaquinol.  
 -!- PATHWAY: Menaquinone biosynthesis; last step.  
 -!- SIMILARITY: Belongs to the ubiE family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: AP005275; BAB97864.1; -;  
 HAMAP: MF\_01813; -; 1.  
 InterPro: IPR001601; Methyltransf.  
 InterPro: IPR000051; SAM bind.  
 InterPro: IPR004033; UbiE/COO5\_MetrF.  
 Pfam: PF01209; UbiE\_methyltran; 1.  
 PROSITE: PS01183; UBI1; FALSE NEG.  
 PROSITE: PS01184; UBI2; FALSE NEG.  
 W Menaquinone biosynthesis; Transferase; Methyltransferase;  
 Complete proteome.  
 Q SEQUENCE 230 AA; 25244 MW; A8548173B7901400 CRC64;  
 -----  
 Query Match 100.0%; Score 19; DB 1; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 1 VAEF 4  
 |||||  
 b 145 VAEF 148  
 -----  
 RESULT 75  
 GIB\_ECOLI STANDARD; PRT; 234 AA.  
 C P24135;  
 T 01-MAR-1992 (Rel. 21, Created)  
 T 01-MAR-1992 (Rel. 21, Last sequence update)  
 T 10-OCT-2003 (Rel. 42, Last annotation update)  
 E Hypothetical protein ygiB.  
 N YGIB OR B3037 OR C3783 OR SF3077 OR S3282.  
 S Escherichia coli.  
 S Escherichia coli O6, and  
 S Shigella flexneri.  
 X Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 X Enterobacteriaceae; Escherichia.  
 X NCBI\_TaxID=562, 217992, 623;  
 UN [1]  
 UN SEQUENCE FROM N.A.  
 IC SPECIES=E. coli;  
 IX MEDLINE=92223101; PubMed=1314093;  
 JA Yang T.-P., Depew R.E.;  
 XT "Nucleotide sequence of a region duplicated in *Escherichia coli* toc  
 mutants";  
 XL Biochim. Biophys. Acta 1130:227-228(1992).  
 [2]  
 UN SEQUENCE FROM N.A.  
 IC SPECIES=E. coli;  
 XX MEDLINE=97426617; PubMed=9276503;  
 XX Blattner F.R.; Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 XX Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 XX Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 XX Mau B., Shao Y.;  
 XT "The complete genome sequence of *Escherichia coli* K-12.";  
 XL Science 277:1453-1474(1997).

[3]  
 RN SEQUENCE FROM N.A.  
 RC SPECIES=E. coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=2238334; PubMed=12471157;  
 RA Weich R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic *Escherichia coli*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 RN [4]  
 RN SEQUENCE FROM N.A.  
 RC SPECIES=S. flexneri; STRAIN=301 / Serotype 2a;  
 RX MEDLINE=2272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qi D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.;  
 RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity  
 through comparison with genomes of *Escherichia coli* K12 and O157.";  
 RL Nucleic Acids Res. 30:4432-4441(2002).  
 RN [5]  
 RN SEQUENCE FROM N.A.  
 RC SPECIES=S. flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 RX MEDLINE=22590274; PubMed=12704152;  
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.;  
 RT "Complete genome sequence and comparative genomics of *Shigella*  
 flexneri serotype 2a strain 2457T.";  
 RL Infect. Immun. 71:2775-2786(2003).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M77129; AAA71875.1; -;  
 DR EMBL: U28377; AAA69205.1; -;  
 DR EMBL: AE000385; AAC76073.1; -;  
 DR EMBL: AE016766; AAN82227.1; -;  
 DR EMBL: AE015319; AAN44555.1; ALT\_INIT.  
 DR EMBL: AE016988; AAP18368.1; -;  
 DR PIR: S22360; S22360.  
 DR Zcogener; EG1164; ygiB.  
 DR Ecogener; EG1164; ygiB.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 234 AA; 24868 MW; 08D5DE931072C966 CRC64;  
 -----  
 Query Match 100.0%; Score 19; DB 1; Length 234;  
 Best Local Similarity 100.0%; Pred. No. 4.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 |||||  
 DB 100 VAEF 103  
 -----  
 Search completed: May 24, 2004, 17:41:04  
 Job time : 8.85714 secs

GenCore version 5.1.6  
Copyright (C) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

un on: May 24, 2004, 17:33:08 ; Search time 28.5714 Seconds  
(without alignments)  
44.173 Million cell updates/sec

itle: US-09-594-978A-2  
effect score: 19  
equene: 1 VAEF 4

coring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

earched: 1017041 seqs, 315518202 residues

otal number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Listing first 1000 summaries

ost-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

atabase :

- 1: SPREMEL 25:\*
- 2: sp\_archaea:\*
- 3: sp\_bacteria:\*
- 4: sp\_fungi:\*
- 5: sp\_human:\*
- 6: sp\_invertebrate:\*
- 7: sp\_mammal:\*
- 8: sp\_mmc:\*
- 9: sp\_organelle:\*
- 10: sp\_phase:\*
- 11: sp\_plant:\*
- 12: sp\_rodent:\*
- 13: sp\_virus:\*
- 14: sp\_vertebrate:\*
- 15: sp\_unclassified:\*
- 16: sp\_rvrius:\*
- 17: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	19	100.0	20	6	Q9T518
2	19	100.0	22	2	Q9R5C0
3	19	100.0	30	5	P82214
4	19	100.0	33	16	O25518
5	19	100.0	35	17	Q8Z219
6	19	100.0	35	6	Q9TOR7
7	19	100.0	44	2	Q48420
8	19	100.0	45	16	Q830H9
9	19	100.0	48	16	O25478
10	19	100.0	51	16	Q7UXL5
11	19	100.0	52	2	Q44941
12	19	100.0	54	16	Q8X3Y8
13	19	100.0	54	16	Q8RIF3
14	19	100.0	61	2	Q8VSE8
15	19	100.0	61	4	Q9P166
16	19	100.0	68	12	Q98GZ9

90	19	100.0	119	5	Q95YG4	Q95YG4 caenorhabdi	163	19	100.0	140	13	Q8JIE5	Q8JIE5 agalychnis
91	19	100.0	119	9	Q853J0	Q853J0 mycobacteri	164	19	100.0	140	16	Q83IC6	Q83IC6 clostridium
92	19	100.0	119	16	Q92ZF9	Q92ZF9 rhizobium m	165	19	100.0	140	16	Q83IC6	Q83IC6 shigella fl
93	19	100.0	119	16	Q87R53	Q87R53 vibrio para	166	19	100.0	142	10	Q23641	Q23641 arabidopsis
94	19	100.0	120	2	Q9ALD4	Q9ALD4 uncultured	167	19	100.0	142	10	Q93368	Q93368 brassica ol
95	19	100.0	120	2	Q9ALH9	Q9ALH9 uncultured	168	19	100.0	142	10	Q41213	Q41213 brassica na
96	19	100.0	120	2	Q9ALB1	Q9ALB1 uncultured	169	19	100.0	142	10	Q42386	Q42386 brassica na
97	19	100.0	120	2	Q31353	Q31353 bacillus ce	170	19	100.0	143	2	Q54526	Q54526 streptomyce
98	19	100.0	120	2	Q9ALH7	Q9ALH7 uncultured	171	19	100.0	143	16	Q8DKS6	Q8DKS6 synchococc
99	19	100.0	120	2	Q9ALH2	Q9ALH2 uncultured	172	19	100.0	143	16	Q82VE3	Q82VE3 nitrosomona
100	19	100.0	120	2	Q9ALE0	Q9ALE0 uncultured	173	19	100.0	144	5	Q24728	Q24728 drosophila
101	19	100.0	120	2	Q9ALD3	Q9ALD3 uncultured	174	19	100.0	144	10	Q9FFG7	Q9FFG7 arabidopsis
102	19	100.0	120	2	Q9ALE0	Q9ALE0 uncultured	175	19	100.0	144	16	Q8ABF9	Q8ABF9 bacteroides
103	19	100.0	121	4	Q722N5	Q722N5 homo sapien	176	19	100.0	145	2	Q54808	Q54808 streptomyce
104	19	100.0	121	9	Q9B066	Q9B066 mycobacteri	177	19	100.0	145	2	Q84127	Q84127 vibrio vuln
105	19	100.0	122	10	Q7XJZ7	Q7XJZ7 oryza sativ	178	19	100.0	145	13	Q9DRE4	Q9DRE4 oncorhynch
106	19	100.0	122	16	Q9PHJ9	Q9PHJ9 xylella fas	179	19	100.0	145	16	Q55965	Q55965 synchocyst
107	19	100.0	123	16	Q9CJM8	Q9CJM8 pasteurella	180	19	100.0	146	5	Q8WTF1	Q8WTF1 ascaris suu
108	19	100.0	123	16	Q9L2J4	Q9L2J4 streptomyce	181	19	100.0	146	5	Q23842	Q23842 drosophila
109	19	100.0	124	9	Q03976	Q03976 bacterioph	182	19	100.0	146	10	Q94AG7	Q94AG7 arabidopsis
110	19	100.0	124	16	Q8FQZ0	Q8FQZ0 corynebacte	183	19	100.0	146	16	Q94AG7	Q94AG7 rhizobium l
111	19	100.0	124	16	Q818K2	Q818K2 bacillus ce	184	19	100.0	146	16	Q8FEX6	Q8FEX6 escherichia
112	19	100.0	125	16	Q89643	Q89643 mycobacteri	185	19	100.0	147	5	Q23929	Q23929 drosophila
113	19	100.0	125	16	Q7TVZ0	Q7TVZ0 mycobacteri	186	19	100.0	147	5	Q24730	Q24730 drosophila
114	19	100.0	125	17	Q26336	Q26336 methanobact	187	19	100.0	147	16	Q69906	Q69906 streptomyce
115	19	100.0	126	2	Q939V9	Q939V9 anabaena so	188	19	100.0	147	17	Q9HPZ8	Q9HPZ8 halobacteri
116	19	100.0	126	6	Q9XSF8	Q9XSF8 bos taurus	189	19	100.0	148	11	Q9CRI8	Q9CRI8 mus musculu
117	19	100.0	126	12	Q84351	Q84351 simian viru	190	19	100.0	148	16	Q9JSS2	Q9JSS2 neisseria m
118	19	100.0	126	16	Q8PN24	Q8PN24 campylobact	191	19	100.0	149	2	Q33KN2	Q33KN2 uncultured
119	19	100.0	126	16	Q7WFX5	Q7WFX5 bordetella	192	19	100.0	149	16	Q8CK75	Q8CK75 rhizobium l
120	19	100.0	127	12	Q91BF3	Q91BF3 spodoptera	193	19	100.0	149	16	Q8Y9M8	Q8Y9M8 listeria mo
121	19	100.0	127	16	Q69714	Q69714 mycobacteri	194	19	100.0	149	16	Q8FRU2	Q8FRU2 corynebacte
122	19	100.0	127	16	Q928Y1	Q928Y1 listeria in	195	19	100.0	149	17	Q8TLY8	Q8TLY8 methanosarc
123	19	100.0	127	16	Q83ET6	Q83ET6 coxiella bu	196	19	100.0	150	16	Q92FE6	Q92FE6 listeria in
124	19	100.0	127	16	Q7TVS6	Q7TVS6 mycobacteri	197	19	100.0	150	16	Q8YAK5	Q8YAK5 listeria mo
125	19	100.0	128	2	Q9S6X8	Q9S6X8 acinetobact	198	19	100.0	151	11	Q9CVJ5	Q9CVJ5 mus musculu
126	19	100.0	128	2	Q83YU7	Q83YU7 acinetobact	199	19	100.0	151	16	Q9ABU9	Q9ABU9 caulobacter
127	19	100.0	128	10	Q9ZSM5	Q9ZSM5 dendrobium	200	19	100.0	152	9	Q8SBU6	Q8SBU6 bacterioph
128	19	100.0	128	16	Q915K0	Q915K0 salmonella	201	19	100.0	152	9	Q7YSS0	Q7YSS0 bacterioph
129	19	100.0	129	10	Q9SBB1	Q9SBB1 arabidopsis	202	19	100.0	153	10	Q9ASA1	Q9ASA1 oryza sativ
130	19	100.0	129	12	Q84339	Q84339 simian viru	203	19	100.0	153	10	Q8S817	Q8S817 oryza sativ
131	19	100.0	130	9	Q8W5Y1	Q8W5Y1 bacterioph	204	19	100.0	153	16	Q8XR05	Q8XR05 ralatonia s
132	19	100.0	130	12	Q8V7C4	Q8V7C4 tt virus. o	205	19	100.0	153	16	Q889K9	Q889K9 lactobacill
133	19	100.0	131	2	Q84CY0	Q84CY0 gamma-prote	206	19	100.0	154	2	Q52066	Q52066 pantoea agg
134	19	100.0	132	2	Q44307	Q44307 anabaena sp	207	19	100.0	154	3	Q86ZD3	Q86ZD3 fusarium sp
135	19	100.0	132	9	Q8SDFO	Q8SDFO pseudomonas	208	19	100.0	155	16	Q7UWB1	Q7UWB1 rhodospirell
136	19	100.0	132	16	Q86418	Q86418 anabaena sp	209	19	100.0	156	16	Q9PMR4	Q9PMR4 campylobact
137	19	100.0	133	2	Q8GH32	Q8GH32 gamma-prote	210	19	100.0	156	17	Q974D5	Q974D5 sulfobobus
138	19	100.0	133	16	Q8RZB3	Q8RZB3 deinooccus	211	19	100.0	157	5	Q9GTU0	Q9GTU0 heterodera
139	19	100.0	133	16	Q8ZGZ5	Q8ZGZ5 versinia pe	212	19	100.0	157	16	Q8FA79	Q8FA79 escherichia
140	19	100.0	133	16	Q8D9V9	Q8D9V9 vibrio vuln	213	19	100.0	157	16	Q8DMU7	Q8DMU7 synchococc
141	19	100.0	133	17	Q8U1Z0	Q8U1Z0 pyrococcus	214	19	100.0	159	12	Q84662	Q84662 paramecium
142	19	100.0	134	2	Q937K7	Q937K7 erwinia chr	215	19	100.0	159	16	Q8FTW6	Q8FTW6 corynebacte
143	19	100.0	134	5	Q96898	Q96898 sacculina c	216	19	100.0	160	2	Q93KL6	Q93KL6 uncultured
144	19	100.0	134	9	Q857W7	Q857W7 mycobacteri	217	19	100.0	161	2	Q99QI3	Q99QI3 shigella fl
145	19	100.0	134	16	Q8ZQT7	Q8ZQT7 salmonella	218	19	100.0	161	2	Q55215	Q55215 streptomyce
146	19	100.0	134	16	Q8Z8C2	Q8Z8C2 salmonella	219	19	100.0	161	2	Q8VSI8	Q8VSI8 shigella fl
147	19	100.0	134	17	Q8Q0F7	Q8Q0F7 methanosarc	220	19	100.0	161	10	Q9AS45	Q9AS45 oryza sativ
148	19	100.0	134	17	Q8Q093	Q8Q093 methanosarc	221	19	100.0	161	10	Q7XAS4	Q7XAS4 oryza sativ
149	19	100.0	135	2	Q939V2	Q939V2 anabaena sp	222	19	100.0	161	16	Q98FZ2	Q98FZ2 rhizobium l
150	19	100.0	135	16	Q986D5	Q986D5 rhizobium l	223	19	100.0	161	16	Q8ZL98	Q8ZL98 salmonella
151	19	100.0	135	16	Q96Z82	Q96Z82 sulfobobus	224	19	100.0	161	16	Q8GZ21	Q8GZ21 bradyrhizob
152	19	100.0	136	16	Q74444	Q74444 synchocyst	225	19	100.0	161	16	Q8Z2B1	Q8Z2B1 salmonella
153	19	100.0	137	2	Q9XBH6	Q9XBH6 bacillus ce	226	19	100.0	162	4	Q8S336	Q8S336 homo sapien
154	19	100.0	137	16	Q9PP94	Q9PP94 campylobact	227	19	100.0	162	11	Q8CQ34	Q8CQ34 mus sp. myb
155	19	100.0	137	16	Q81X11	Q81X11 bacillus an	228	19	100.0	163	4	Q86WJ3	Q86WJ3 homo sapien
156	19	100.0	137	16	Q914R9	Q914R9 bacillus ce	229	19	100.0	163	13	Q7ZXF9	Q7ZXF9 xenopus lae
157	19	100.0	138	16	Q9KQ08	Q9KQ08 vibrio chol	230	19	100.0	163	16	Q07237	Q07237 mycobacteri
158	19	100.0	138	16	Q9KXN0	Q9KXN0 bradyrhizob	231	19	100.0	163	16	Q7UZA7	Q7UZA7 pyrobaculum
159	19	100.0	139	16	Q8K9P4	Q8K9P4 bacillus ha	232	19	100.0	163	17	Q8ZX97	Q8ZX97 pyrobaculum
160	19	100.0	139	16	Q8D6C8	Q8D6C8 vibrio vuln	233	19	100.0	164	16	Q8Y7N8	Q8Y7N8 listeria mo
161	19	100.0	139	16	Q8ZNK0	Q8ZNK0 streptomyce	234	19	100.0	164	16	Q9KZL0	Q9KZL0 streptomyce
162	19	100.0	140	11	Q9ESY3	Q9ESY3 mus musculu	235	19	100.0	164	17	Q29902	Q29902 archaeoglob

236 19 100.0 165 4 Q86ED1  
 237 19 100.0 165 5 Q8SRD1  
 238 19 100.0 165 11 Q9D6S8  
 239 19 100.0 166 2 Q52661  
 240 19 100.0 166 10 Q8R296  
 241 19 100.0 166 10 Q948R0  
 242 19 100.0 166 17 Q8ZVJ9  
 243 19 100.0 167 2 Q8GNC0  
 244 19 100.0 168 2 Q47850  
 245 19 100.0 168 16 Q8ZFV0  
 246 19 100.0 169 16 Q8X8N2  
 247 19 100.0 169 16 Q7UKV3  
 248 19 100.0 169 16 Q7U5D8  
 249 19 100.0 170 16 Q83JC6  
 250 19 100.0 170 16 Q881F4  
 251 19 100.0 171 10 Q84QL9  
 252 19 100.0 171 12 Q8QJL5  
 253 19 100.0 172 10 Q9AWP5  
 254 19 100.0 172 11 Q85179  
 255 19 100.0 172 11 Q8K164  
 256 19 100.0 172 11 Q85177  
 257 19 100.0 172 17 Q9UZ36  
 258 19 100.0 172 17 Q8U1U0  
 259 19 100.0 173 5 Q20526  
 260 19 100.0 173 5 Q8S8X9  
 261 19 100.0 174 11 Q8CIG2  
 262 19 100.0 174 16 Q8ECL6  
 263 19 100.0 174 16 Q33289  
 264 19 100.0 174 16 Q7V6A9  
 265 19 100.0 174 16 Q7TX22  
 266 19 100.0 175 11 Q925A9  
 267 19 100.0 175 16 Q8A2B2  
 268 19 100.0 175 16 Q82DE1  
 269 19 100.0 175 17 Q58465  
 270 19 100.0 176 16 Q7VD73  
 271 19 100.0 176 16 Q976D7  
 272 19 100.0 178 5 Q8MTA3  
 273 19 100.0 178 16 Q7VJ34  
 274 19 100.0 178 16 Q7UTJ4  
 275 19 100.0 179 2 Q54157  
 276 19 100.0 179 11 Q8R057  
 277 19 100.0 179 16 Q7VKQ4  
 278 19 100.0 180 2 Q8K5U7  
 279 19 100.0 180 2 Q8K5U7  
 280 19 100.0 180 10 Q8EXU1  
 281 19 100.0 180 10 Q9M098  
 282 19 100.0 180 10 Q8LXD3  
 283 19 100.0 180 16 Q8ZPH9  
 284 19 100.0 180 16 Q8Z707  
 285 19 100.0 180 16 Q8J330  
 286 19 100.0 180 16 Q8ZEC1  
 287 19 100.0 180 16 Q81DE6  
 288 19 100.0 180 16 Q05559  
 289 19 100.0 180 16 Q7UAE1  
 290 19 100.0 181 5 Q8MU99  
 291 19 100.0 181 16 Q8YSB5  
 292 19 100.0 182 16 Q7V2G9  
 293 19 100.0 182 17 Q8TWT2  
 294 19 100.0 183 4 Q8TCQ9  
 295 19 100.0 184 16 Q8X5C1  
 296 19 100.0 184 16 Q92LY9  
 297 19 100.0 184 16 Q87YT2  
 298 19 100.0 184 16 Q7VC97  
 299 19 100.0 185 16 Q8UHU5  
 300 19 100.0 185 16 Q9RZ9H  
 301 19 100.0 186 2 Q56745  
 302 19 100.0 186 2 P72209  
 303 19 100.0 187 2 Q93P57  
 304 19 100.0 187 16 Q9AFD5  
 305 19 100.0 187 16 Q9ZWB0  
 306 19 100.0 188 5 Q8T3L7  
 307 19 100.0 188 16 Q83722  
 308 19 100.0 188 16 Q8ZH37

309 19 100.0 188 16 Q8XC46  
 310 19 100.0 188 17 Q8TNU0  
 311 19 100.0 189 2 Q84CM0  
 312 19 100.0 189 5 Q20199  
 313 19 100.0 189 8 Q8WJB4  
 314 19 100.0 189 10 Q07299  
 315 19 100.0 189 16 Q8U549  
 316 19 100.0 189 16 Q8FC12  
 317 19 100.0 190 4 Q86F12  
 318 19 100.0 190 16 Q93RS6  
 319 19 100.0 191 2 Q93KL3  
 320 19 100.0 191 10 Q94LF2  
 321 19 100.0 191 10 Q84MM8  
 322 19 100.0 191 16 Q9KKV1  
 323 19 100.0 191 16 Q8EAK2  
 324 19 100.0 192 10 Q8L647  
 325 19 100.0 192 16 Q8UG64  
 326 19 100.0 192 16 Q9ADP7  
 327 19 100.0 192 16 Q83D37  
 328 19 100.0 193 2 P74983  
 329 19 100.0 193 16 Q9KY43  
 330 19 100.0 193 16 Q82B77  
 331 19 100.0 194 2 Q9EVD1  
 332 19 100.0 194 2 Q8Q8N2  
 333 19 100.0 194 2 Q8GE12  
 334 19 100.0 194 2 Q83XJ2  
 335 19 100.0 194 2 Q7X3V2  
 336 19 100.0 194 16 Q9A7S1  
 337 19 100.0 194 16 Q8UBS8  
 338 19 100.0 195 10 Q9LMI9  
 339 19 100.0 195 16 Q89TU4  
 340 19 100.0 195 16 Q88SG8  
 341 19 100.0 196 2 Q93KM6  
 342 19 100.0 196 5 Q9GNR5  
 343 19 100.0 196 5 Q93568  
 344 19 100.0 196 5 Q9GNR7  
 345 19 100.0 196 11 Q9CVX4  
 346 19 100.0 196 16 Q8FT59  
 347 19 100.0 196 16 Q8DGL7  
 348 19 100.0 196 16 Q8Z2K7  
 349 19 100.0 197 2 Q93KN1  
 350 19 100.0 197 9 Q8H9Q1  
 351 19 100.0 197 16 Q9CGV5  
 352 19 100.0 197 17 Q8TZN4  
 353 19 100.0 198 2 Q8KRL5  
 354 19 100.0 198 16 Q98NC6  
 355 19 100.0 198 16 Q50927  
 356 19 100.0 199 4 Q9H3L5  
 357 19 100.0 199 4 Q8EX83  
 358 19 100.0 199 11 Q8BXC6  
 359 19 100.0 199 11 Q8BUJ3  
 360 19 100.0 199 16 Q8XTM5  
 361 19 100.0 199 16 Q7U983  
 362 19 100.0 200 16 Q8ZMY2  
 363 19 100.0 200 16 Q910A5  
 364 19 100.0 200 16 Q8ZQA6  
 365 19 100.0 200 16 Q8Z7W8  
 366 19 100.0 201 2 Q9X407  
 367 19 100.0 201 10 Q48848  
 368 19 100.0 201 16 Q88I27  
 369 19 100.0 202 2 Q8RNS6  
 370 19 100.0 202 16 Q9Y8Y2  
 371 19 100.0 202 16 Q69982  
 372 19 100.0 202 16 P73372  
 373 19 100.0 203 2 Q9RBB8  
 374 19 100.0 203 10 Q49A7  
 375 19 100.0 203 16 Q8PB95  
 376 19 100.0 203 16 Q98Y88  
 377 19 100.0 203 16 Q88DQ5  
 378 19 100.0 204 2 Q9R8K6  
 379 19 100.0 204 2 Q83XB6  
 380 19 100.0 204 4 Q9NZ83  
 381 19 100.0 204 4 Q7Z2N7

Q8XC46 escherichia  
 Q8tnj0 methanosarc  
 Q8cm0 streptomyce  
 Q20199 caenorhabdi  
 Q8WJB4 eriosorus h  
 Q07299 zea mays (m  
 Q8U549 agrobacteri  
 Q8fc12 escherichia  
 Q96f12 homo sapien  
 Q93rs6 streptomyce  
 Q93kl3 uncultured  
 Q94lf2 oryza sativ  
 Q84mw8 oryza sativ  
 Q9kkv1 vibrio chol  
 Q8eak2 shewanella  
 Q8l647 deschampsia  
 Q9adp7 agrobacteri  
 Q83d37 streptomyce  
 Q83d37 coxiella bu  
 P74983 yersinia en  
 Q9ky43 streptomyce  
 Q82b77 streptomyce  
 Q9evd1 erwinia amy  
 Q8ggn2 serratia md  
 Q8ge12 erwinia pyr  
 Q83xj2 vibrio angu  
 Q7x3v2 erwinia sp.  
 Q9a7s1 caulobacter  
 Q8ub8 agrobacteri  
 Q9lmi9 arabidopsis  
 Q89tu4 bradyrhizob  
 Q88sg8 lactobacill  
 Q93km6 uncultured  
 Q9gnr5 stylonychia  
 Q93568 caenorhabdi  
 Q9gnr7 stylonychia  
 Q9cvx4 mus musculu  
 Q8ft59 corynebacte  
 Q8dgl7 synechococc  
 Q8z2k7 salmonella  
 Q93kn1 uncultured  
 Q8h9q1 vibrio harv  
 Q9cgv5 lactococcus  
 Q8tzn4 pyrococcus  
 Q8xrl5 rhodocycclus  
 Q98nc6 rhizobium 1  
 Q50927 nitrosomona  
 Q9h3l5 homo sapien  
 Q85x83 homo sapien  
 Q8bxc6 mus musculu  
 Q8bjj3 mus musculu  
 Q8xtm5 talstonia s  
 Q7u983 synechococc  
 Q8zmy2 salmonella  
 Q910a5 streptomyce  
 Q8zqa6 salmonella  
 Q8z7w8 salmonella  
 Q9x407 methylsulfi  
 Q48848 arabidopsis  
 Q88i27 rhizobium 1  
 Q98y82 listeria mo  
 Q69982 streptomyce  
 P73372 synechocyst  
 Q9rbb8 arthrobacte  
 Q49a7 oryza sativ  
 Q8pb95 xanthomonas  
 Q98y88 clostridium  
 Q88dg5 pseudomonas  
 Q9rbk6 xanthomonas  
 Q83xe6 xanthomonas  
 Q9nz83 homo sapien  
 Q7z2n7 homo sapien

382	19	100.0	204	16	Q92F26	Q92f26 rickettsia	455	19	100.0	224	16	Q8NRP2	Q8nrp2 corynebacte
383	19	100.0	204	16	Q8PQC4	Q8pqc4 xanthomonas	456	19	100.0	225	16	Q986L5	Q986l5 rhizobium 1
384	19	100.0	204	16	Q93J78	Q93j78 streptomyc	457	19	100.0	225	16	Q825R9	Q825r9 streptomyc
385	19	100.0	204	16	Q7U559	Q7u559 shigella fl	458	19	100.0	226	2	Q9RBD7	Q9rbd7 arthrobacte
386	19	100.0	205	16	Q8X810	Q8x810 escherichia	459	19	100.0	226	2	Q84BB9	Q84bb9 streptomyc
387	19	100.0	205	16	Q8GSL8	Q8gsl8 bifidobacte	460	19	100.0	226	5	Q8WPD3	Q8wpd3 paracentrot
388	19	100.0	205	16	Q8F810	Q8f810 escherichia	461	19	100.0	226	11	Q8C5E1	Q8c5e1 mus musculu
389	19	100.0	205	16	Q8F810	Q8f810 escherichia	462	19	100.0	226	16	Q8YDM2	Q8ydm2 bruceella su
390	19	100.0	205	16	Q8DVM4	Q8dvm4 streptococc	463	19	100.0	226	16	Q8FUS4	Q8fus4 bruceella su
391	19	100.0	206	2	Q831V9	Q831v9 shigella fl	464	19	100.0	227	2	Q9KJ22	Q9kjj2 pseudomonas
392	19	100.0	206	11	Q8C0Y5	Q8c0y5 mus musculu	465	19	100.0	227	10	Q8GRV7	Q8grv7 oryza sativ
393	19	100.0	206	16	Q8EC34	Q8ec34 shewanella	466	19	100.0	227	17	Q8ZVD4	Q8zvd4 pyrobaculum
394	19	100.0	207	10	Q8H4R2	Q8h4r2 oryza sativ	467	19	100.0	228	16	Q8Z6Q3	Q8z6q3 salmonella
395	19	100.0	208	16	Q8PD60	Q8pd60 xanthomonas	468	19	100.0	228	16	Q8KS24	Q8ks24 deinococcus
396	19	100.0	208	16	Q82AM2	Q82am2 streptomyc	469	19	100.0	228	16	Q8R153	Q8r153 yersinia pe
397	19	100.0	209	2	Q9RBB6	Q9rb6 arthrobacte	470	19	100.0	228	16	Q8FSE2	Q8fse2 corynebacte
398	19	100.0	209	4	Q9BYW9	Q9byw9 homo sapien	471	19	100.0	228	16	Q894X0	Q894x0 clostridium
399	19	100.0	209	16	Q8UF55	Q8uf55 agrobacteri	472	19	100.0	228	16	Q894X0	Q894x0 clostridium
400	19	100.0	210	16	Q8ZPD7	Q8zpd7 salmonella	473	19	100.0	229	10	Q9CAS5	Q9cas5 arabidopsis
401	19	100.0	210	16	Q8G4Z4	Q8g4z4 bifidobacte	474	19	100.0	229	16	Q66753	Q66753 aquifex aeo
402	19	100.0	210	16	Q83190	Q83190 shigella fl	475	19	100.0	229	16	Q8GSD8	Q8gsd8 staphylococ
403	19	100.0	211	10	Q7XA92	Q7xa92 fragaria an	476	19	100.0	229	16	Q832X6	Q832x6 enterococcu
404	19	100.0	211	16	Q92R80	Q92r80 rhizobium m	477	19	100.0	230	2	Q9RBB0	Q9rb0 arthrobacte
405	19	100.0	211	16	Q8P940	Q8p940 xanthomonas	478	19	100.0	230	5	Q76405	Q76405 caenorhabdi
406	19	100.0	212	2	Q9RBB7	Q9rb7 arthrobacte	479	19	100.0	230	10	Q84RH7	Q84rh7 lycopersico
407	19	100.0	212	10	Q8RUH2	Q8ruh2 oryza sativ	480	19	100.0	230	10	Q8UHL0	Q8uhl0 agrobacteri
408	19	100.0	212	16	Q914X9	Q914x9 pseudomonas	481	19	100.0	230	16	Q8DOF9	Q8dof9 yersinia pe
409	19	100.0	213	10	Q8H3F6	Q8h3f6 oryza sativ	482	19	100.0	230	16	Q7WQY2	Q7wgy2 bordetella
410	19	100.0	213	16	Q9KP59	Q9kp59 vibrio chol	483	19	100.0	230	16	Q7W204	Q7w204 bordetella
411	19	100.0	213	16	Q8YKX0	Q8ykx0 anabaena sp	484	19	100.0	230	16	Q7VUL1	Q7vul1 bordetella
412	19	100.0	213	16	Q8NTZ8	Q8ntz8 corynebacte	485	19	100.0	231	5	Q8MXA3	Q8mxa3 trypanosoma
413	19	100.0	213	16	Q87VGS	Q87vg8 pseudomonas	486	19	100.0	231	5	Q81684	Q81684 drosophila
414	19	100.0	213	17	Q9HR09	Q9hr09 halobacteri	487	19	100.0	231	16	Q9HXR1	Q9hxr1 pseudomonas
415	19	100.0	214	2	Q84380	Q84380 lactococcus	488	19	100.0	231	16	Q92T12	Q92t12 rhizobium 1
416	19	100.0	214	2	Q8KP00	Q8kp00 gamma-prote	489	19	100.0	231	16	Q92E94	Q92e94 listeria in
417	19	100.0	214	2	Q9RC84	Q9rc84 bacillus ha	490	19	100.0	231	16	Q8YXG7	Q8yxx7 bruceella me
418	19	100.0	215	2	Q92480	Q92480 agrobacteri	491	19	100.0	231	16	Q8Y9H1	Q8y9h1 listeria mo
419	19	100.0	215	10	Q22136	Q22136 arabidopsis	492	19	100.0	231	16	Q8XUT7	Q8xut7 raietonia s
420	19	100.0	215	16	Q8PDH0	Q8pdh0 rhizobium 1	493	19	100.0	231	16	Q8FOL3	Q8fgl3 corynebacte
421	19	100.0	215	16	Q8PQ35	Q8pq35 xanthomonas	494	19	100.0	232	2	Q9RSD1	Q9rbd1 arthrobacte
422	19	100.0	216	2	Q9RBD2	Q9rb2 arthrobacte	495	19	100.0	232	16	Q9A6A4	Q9a6a4 caulobacter
423	19	100.0	216	14	P97979	P97979 unidentified	496	19	100.0	232	16	P96241	P96241 mycobacteri
424	19	100.0	216	16	Q92QS1	Q92qs1 rhizobium m	497	19	100.0	232	16	Q7TVJ7	Q7tvj7 mycobacteri
425	19	100.0	216	16	Q8NVA0	Q8nva0 pseudomonas	498	19	100.0	232	17	Q8PVL0	Q8pvl0 methanosarc
426	19	100.0	216	16	Q8GGA2	Q8gga2 pseudomonas	499	19	100.0	232	2	Q9RBC9	Q9rb9 arthrobacte
427	19	100.0	217	17	Q8UIA8	Q8uia8 pyrococcus	500	19	100.0	234	5	Q9N327	Q9n327 caenorhabdi
428	19	100.0	218	2	Q9REK4	Q9rek4 arthrobacte	501	19	100.0	234	10	Q9FQC4	Q9fgc4 zea mays (m
429	19	100.0	218	10	Q40630	Q40630 oryza sativ	502	19	100.0	234	10	Q9AKL2	Q9akl2 arabidopsis
430	19	100.0	218	10	Q9L2V5	Q9l2v5 arabidopsis	503	19	100.0	234	10	Q9XIF8	Q9xif8 arabidopsis
431	19	100.0	218	16	Q91470	Q91470 pseudomonas	504	19	100.0	234	16	Q8YSQ0	Q8ysq0 anabaena sp
432	19	100.0	218	16	Q8ZES1	Q8zes1 yersinia pe	505	19	100.0	234	16	Q8XBP4	Q8xbp4 escherichia
433	19	100.0	218	16	Q88NB6	Q88nb6 pseudomonas	506	19	100.0	234	16	Q81D13	Q81d13 bacillus ce
434	19	100.0	219	16	Q82E71	Q82e71 listeria in	507	19	100.0	235	2	Q9AKL9	Q9akl9 rickettsia
435	19	100.0	219	16	Q69984	Q69984 streptomyc	508	19	100.0	235	2	Q9AKB8	Q9akb8 rickettsia
436	19	100.0	219	16	Q82KE5	Q82ke5 streptomyc	509	19	100.0	235	2	Q9AKG3	Q9akg3 rickettsia
437	19	100.0	219	16	Q7UGG1	Q7ugg1 rhodospirell	510	19	100.0	235	2	Q849C9	Q849c9 streptomyc
438	19	100.0	219	17	Q26842	Q26842 methanobact	511	19	100.0	235	9	Q8W6X8	Q8w6x8 cyanophage
439	19	100.0	220	17	Q9HQX1	Q9hqx1 halobacteri	512	19	100.0	235	16	Q92XB1	Q92xb1 rhizobium m
440	19	100.0	221	4	Q96GF5	Q96gf5 homo sapien	513	19	100.0	235	16	Q921J4	Q921j4 rickettsia
441	19	100.0	221	16	Q9KAB1	Q9kab1 bacillus ha	514	19	100.0	235	16	Q8XYZ8	Q8xyz8 rhodocyclu
442	19	100.0	221	16	Q93J34	Q93j34 streptomyc	515	19	100.0	236	2	Q9JPA0	Q9jpa0 rhodocyclu
443	19	100.0	221	16	Q82CX2	Q82cx2 streptomyc	516	19	100.0	236	10	Q91VA8	Q91va8 arabidopsis
444	19	100.0	222	16	Q9RBD9	Q9rb9 arthrobacte	517	19	100.0	236	10	Q9SK05	Q9sk05 arabidopsis
445	19	100.0	222	16	Q8F9P8	Q8f9p8 leptospira	518	19	100.0	237	16	Q8UBG5	Q8ubg5 agrobacteri
446	19	100.0	223	5	Q8SUL0	Q8sul0 encephalito	519	19	100.0	237	2	Q9RBC0	Q9rb0 arthrobacte
447	19	100.0	223	10	Q94B57	Q94b57 arabidopsis	520	19	100.0	237	16	Q9KFS9	Q9kfs9 bacillus ha
448	19	100.0	223	16	Q9ZBP8	Q9zbp8 streptomyc	521	19	100.0	237	16	Q9N57	Q9n57 rhizobium 1
449	19	100.0	223	16	Q883K9	Q88k9 pseudomonas	522	19	100.0	237	16	Q91V4	Q91v4 clostridium
450	19	100.0	223	16	Q8XG51	Q8xg51 salmonella	523	19	100.0	237	16	Q8ELK7	Q8elk7 oceanobacil
451	19	100.0	224	3	Q03362	Q03362 saccharomyc	524	19	100.0	237	16	Q8TLK7	Q8tlk7 oceanobacil
452	19	100.0	224	4	Q92799	Q92799 homo sapien	525	19	100.0	237	17	Q8TMA2	Q8tma2 methanosarc
453	19	100.0	224	11	Q99MG0	Q99mg0 rattus norv	526	19	100.0	238	2	Q9RBP4	Q9rbp4 rhodococcus
454	19	100.0	224	16	Q8Y8J7	Q8y8j7 listeria mo	527	19	100.0	238	16	Q9CIV5	Q9civ5 lactococcus

528	19	100.0	238	16	Q9L245	Q9L246 streptomyc	601	19	100.0	253	16	Q8X4N5	Q8X4N5 escherichia
529	19	100.0	239	3	Q9UVW7	Q9UVW7 candida alb	602	19	100.0	253	16	Q9AKO7	Q9AKO7 streptomyc
530	19	100.0	239	16	Q836K1	Q836K1 nitrosococcu	603	19	100.0	253	16	Q8F9B8	Q8F9B8 escherichia
531	19	100.0	239	16	Q8ZUL5	Q8ZUL5 nitrosococcu	604	19	100.0	253	16	Q83QB3	Q83QB3 escherichia fl
532	19	100.0	239	17	Q9V213	Q9V213 pyrococcus	605	19	100.0	253	17	Q9YB95	Q9YB95 aeropyrum p
533	19	100.0	240	17	Q30144	Q30144 archaeoglob	606	19	100.0	254	2	Q9REK0	Q9REK0 athrobacte
534	19	100.0	241	2	Q7WTF7	Q7WTF7 streptomyc	607	19	100.0	254	2	Q9Z5Q1	Q9Z5Q1 clostridium
535	19	100.0	241	2	Q9SVW3	Q9SVW3 arabisopsis	608	19	100.0	254	8	Q98RW7	Q98RW7 guillardia
536	19	100.0	241	10	Q7XR82	Q7XR82 oryza sativ	609	19	100.0	254	10	Q94A55	Q94A55 arabisopsis
537	19	100.0	241	16	Q7WCN5	Q7WCN5 bordetella	610	19	100.0	254	10	Q22536	Q22536 oryza sativ
538	19	100.0	241	16	Q7WA18	Q7WA18 bordetella	611	19	100.0	254	10	Q9ATM0	Q9ATM0 zea mays (m
539	19	100.0	242	5	Q81683	Q81683 drosophila	612	19	100.0	254	16	Q8ZGH5	Q8ZGH5 yersinia pe
540	19	100.0	242	5	Q81682	Q81682 drosophila	613	19	100.0	255	2	Q9RBE0	Q9RBE0 athrobacte
541	19	100.0	242	16	Q8Y0T0	Q8Y0T0 ralstonia s	614	19	100.0	255	8	Q950V2	Q950V2 tetrahymena
542	19	100.0	242	16	Q8DH96	Q8DH96 synchococc	615	19	100.0	255	16	Q89JK6	Q89JK6 bradyrhizob
543	19	100.0	242	16	Q8CY98	Q8CY98 streptococc	616	19	100.0	256	2	Q9RBD5	Q9RBD5 athrobacte
544	19	100.0	243	2	Q9AN57	Q9AN57 bradyrhizob	617	19	100.0	256	8	Q9T7M4	Q9T7M4 tetrahymena
545	19	100.0	243	16	Q8GOY5	Q8GOY5 brucella su	618	19	100.0	256	10	Q94HB4	Q94HB4 oryza sativ
546	19	100.0	244	6	Q8HXL4	Q8HXL4 sus scrofa	619	19	100.0	256	10	Q65325	Q65325 arabisopsis
547	19	100.0	244	16	Q8DSN5	Q8DSN5 streptococc	620	19	100.0	256	10	Q7XF17	Q7XF17 oryza sativ
548	19	100.0	244	16	Q7UQX9	Q7UQX9 rhodopirell	621	19	100.0	256	12	Q9WN21	Q9WN21 frangipani
549	19	100.0	245	10	Q940J5	Q940J5 arabisopsis	622	19	100.0	257	10	Q8GUK5	Q8GUK5 arabisopsis
550	19	100.0	245	10	Q7XK61	Q7XK61 oryza sativ	623	19	100.0	257	11	Q9CP85	Q9CP85 mus musculu
551	19	100.0	246	6	Q28535	Q28535 mustela vis	624	19	100.0	257	11	Q99UB5	Q99UB5 mus musculu
552	19	100.0	246	16	Q8PFI51	Q8PFI51 xanthomonas	625	19	100.0	257	16	Q92102	Q92102 rhizobium m
553	19	100.0	246	16	Q8P6V1	Q8P6V1 xanthomonas	626	19	100.0	257	16	Q8NLT7	Q8NLT7 streptococ
554	19	100.0	246	16	Q9RDB0	Q9RDB0 streptomyc	627	19	100.0	257	16	Q9X7Q4	Q9X7Q4 streptomyc
555	19	100.0	246	16	Q8FMA7	Q8FMA7 corynebacte	628	19	100.0	258	2	Q9RBB1	Q9RBB1 athrobacte
556	19	100.0	246	16	Q89MY3	Q89MY3 bradyrhizob	629	19	100.0	258	13	Q7S2S8	Q7S2S8 xenopus lae
557	19	100.0	246	16	Q8ZAC3	Q8ZAC3 streptomyc	630	19	100.0	258	16	Q9HZN9	Q9HZN9 pseudomonas
558	19	100.0	247	2	Q48830	Q48830 lactobacill	631	19	100.0	258	16	Q9XBL2	Q9XBL2 streptomyc
559	19	100.0	247	16	Q88ZC3	Q88ZC3 lactobacill	632	19	100.0	258	16	Q7WCV7	Q7WCV7 bordetella
560	19	100.0	248	5	Q97008	Q97008 leishmania	633	19	100.0	258	16	Q7WSC3	Q7WSC3 bordetella
561	19	100.0	248	10	Q94B99	Q94B99 oryza sativ	634	19	100.0	258	16	Q7VZN8	Q7VZN8 bordetella
562	19	100.0	248	10	Q39935	Q39935 helianthus	635	19	100.0	258	17	Q9HLA1	Q9HLA1 thermoplas
563	19	100.0	248	10	Q81216	Q81216 zea mays (m	636	19	100.0	259	2	Q338F2	Q338F2 eubacterium
564	19	100.0	248	10	Q84RL6	Q84RL6 zea mays (m	637	19	100.0	259	2	Q9VMK6	Q9VMK6 pseudomonas
565	19	100.0	248	10	Q7XAG1	Q7XAG1 oryza sativ	638	19	100.0	259	13	Q7ZTI9	Q7ZTI9 xenopus lae
566	19	100.0	248	16	Q8ZG59	Q8ZG59 yersinia pe	639	19	100.0	259	16	F74082	F74082 synchocyst
567	19	100.0	248	16	Q8A971	Q8A971 bacteroides	640	19	100.0	259	16	Q9PJ15	Q9PJ15 campylobact
568	19	100.0	248	17	Q9YCF6	Q9YCF6 aeropyrum p	641	19	100.0	259	16	Q98X61	Q98X61 rhizobium l
569	19	100.0	249	5	Q62042	Q62042 caenorhabdi	642	19	100.0	259	16	Q7U9R9	Q7U9R9 synchococc
570	19	100.0	249	10	Q9LXK3	Q9LXK3 hordeum vul	643	19	100.0	259	17	Q8ZV71	Q8ZV71 pyrobaculum
571	19	100.0	249	10	Q9ATU9	Q9ATU9 zea mays (m	644	19	100.0	260	2	Q9RBA7	Q9RBA7 athrobacte
572	19	100.0	249	10	Q7XU33	Q7XU33 oryza sativ	645	19	100.0	260	2	Q8G976	Q8G976 pseudomonas
573	19	100.0	249	12	Q91914	Q91914 culex nigri	646	19	100.0	260	2	Q849Q8	Q849Q8 pseudomonas
574	19	100.0	249	16	Q9PKD5	Q9PKD5 chlamydia m	647	19	100.0	260	16	Q8YVJ0	Q8YVJ0 anabaena sp
575	19	100.0	249	16	Q8PG49	Q8PG49 xanthomonas	648	19	100.0	260	16	Q8F545	Q8F545 corynebacte
576	19	100.0	249	16	Q8F1P0	Q8F1P0 leptospira	649	19	100.0	260	16	Q813A0	Q813A0 bacillus ce
577	19	100.0	249	16	Q33N41	Q33N41 coxiella bu	650	19	100.0	260	16	Q7WN67	Q7WN67 bordetella
578	19	100.0	249	16	Q82NM5	Q82NM5 streptomyc	651	19	100.0	260	16	Q7WU07	Q7WU07 bordetella
579	19	100.0	249	17	Q29512	Q29512 archaeoglob	652	19	100.0	260	16	Q7VW38	Q7VW38 bordetella
580	19	100.0	250	10	Q9AT18	Q9AT18 zea mays (m	653	19	100.0	261	2	Q66247	Q66247 actinobacil
581	19	100.0	250	10	Q84JP6	Q84JP6 arabisopsis	654	19	100.0	261	3	Q9P8F1	Q9P8F1 glomus moss
582	19	100.0	250	11	Q8C4H5	Q8C4H5 mus musculu	655	19	100.0	261	3	Q9P8B9	Q9P8B9 glomus moss
583	19	100.0	250	16	Q25228	Q25228 helicobacte	656	19	100.0	261	3	Q9P8B9	Q9P8B9 glomus moss
584	19	100.0	251	2	Q33A69	Q33A69 gamma-prote	657	19	100.0	261	6	Q95LNG	Q95LNG macaca fasc
585	19	100.0	251	5	Q8SVK8	Q8SVK8 encephalito	658	19	100.0	261	11	Q8K2L3	Q8K2L3 mus musculu
586	19	100.0	251	11	Q9D4Z4	Q9D4Z4 mus musculu	659	19	100.0	261	11	Q9DCE1	Q9DCE1 mus musculu
587	19	100.0	251	17	Q28320	Q28320 archaeoglob	660	19	100.0	261	11	Q9JWJ6	Q9JWJ6 mus musculu
588	19	100.0	252	2	Q03540	Q03540 escherichia	661	19	100.0	261	11	Q8BH40	Q8BH40 mus musculu
589	19	100.0	252	9	Q8HA20	Q8HA20 bacterioph	662	19	100.0	261	16	Q9K9M5	Q9K9M5 bacillus ha
590	19	100.0	252	10	Q94CS9	Q94CS9 oryza sativ	663	19	100.0	261	16	Q987X7	Q987X7 rhizobium l
591	19	100.0	252	10	Q7XA62	Q7XA62 oryza sativ	664	19	100.0	261	16	Q8UCA1	Q8UCA1 agrobacteri
592	19	100.0	252	16	Q9A534	Q9A534 caulobacter	665	19	100.0	261	16	Q8U9J3	Q8U9J3 agrobacteri
593	19	100.0	252	16	Q88U86	Q88U86 lactobacill	666	19	100.0	261	16	Q87IH4	Q87IH4 pseudomonas
594	19	100.0	252	16	Q7UT61	Q7UT61 rhodopirell	667	19	100.0	261	16	Q7WPF1	Q7WPF1 bordetella
595	19	100.0	252	17	Q97TV6	Q97TV6 pectolobus	668	19	100.0	261	16	Q7WBP9	Q7WBP9 bordetella
596	19	100.0	253	2	Q8KH15	Q8KH15 pectobacter	669	19	100.0	262	2	Q8RSJ2	Q8RSJ2 gamma-prote
597	19	100.0	253	5	Q9SU55	Q9SU55 drosophila	670	19	100.0	262	3	Q9P8B8	Q9P8B8 glomus moss
598	19	100.0	253	10	Q8W371	Q8W371 oryza sativ	671	19	100.0	262	3	Q9WZP2	Q9WZP2 thermotoga
599	19	100.0	253	10	Q7XD92	Q7XD92 oryza sativ	672	19	100.0	262	16	Q8EDY1	Q8EDY1 shewanella
600	19	100.0	253	16	Q8ZFH9	Q8ZFH9 yersinia pe	673	19	100.0	262	17	Q8U0A9	Q8U0A9 pyrococcus

674	19	100.0	263	16	Q8D064	Q8d064 yersinia pe	747	19	100.0	276	16	Q8Z016	Q8z016 anabaena sp
675	19	100.0	263	16	Q8A005	Q8a0g5 bacteroides	748	19	100.0	276	16	Q8PF23	Q8ppz3 pseudomonas
676	19	100.0	263	16	Q7Wn22	Q7wn22 bordetella	749	19	100.0	276	16	Q8A1P0	Q8a1p0 pseudomonas
677	19	100.0	263	16	Q7WbJ9	Q7wbj9 bordetella	750	19	100.0	277	2	Q8S287	Q8s287 sphingomona
678	19	100.0	264	4	Q8TDN7	Q8tdn7 homo sapien	751	19	100.0	277	2	P72544	P72544 synchococc
679	19	100.0	264	10	Q940D9	Q940d9 pinus taeda	752	19	100.0	277	16	Q8P8X6	Q8p8x6 neisseria m
680	19	100.0	265	4	Q8TDC1	Q8tdc1 homo sapien	753	19	100.0	278	16	Q9JXZ5	Q9jxz5 neisseria m
681	19	100.0	265	5	Q15953	Q15953 drosophila	754	19	100.0	278	16	Q9JXZ5	Q9jxz5 neisseria m
682	19	100.0	265	10	Q8S3T6	Q8s3t6 oryza sativ	755	19	100.0	278	16	Q8S3T6	Q8s3t6 oryza sativ
683	19	100.0	265	11	Q91XZ8	Q91xz8 rattus norv	756	19	100.0	278	16	Q8S694	Q8s694 treponema p
684	19	100.0	265	17	Q978W2	Q978w2 thermoplas	757	19	100.0	279	16	Q8E5D6	Q8e5d6 streptococ
685	19	100.0	267	2	Q05364	Q05364 actinobacil	758	19	100.0	279	16	Q9XK81	Q9xk81 rhizobium l
686	19	100.0	267	2	Q83Y72	Q83y72 gamma-prote	759	19	100.0	279	16	Q7WCY6	Q7wcy6 bordetella
687	19	100.0	267	10	Q8GZQ2	Q8gzq2 zea mays (m	760	19	100.0	279	17	Q7VTG9	Q7vtg9 bordetella
688	19	100.0	267	11	Q8Bw25	Q8bw25 mus musculu	761	19	100.0	280	2	Q9HSE9	Q9haf9 halobacteri
689	19	100.0	267	16	Q8A9V3	Q8a9v3 caulobacter	762	19	100.0	280	2	P71101	P71101 curtobacter
690	19	100.0	267	16	Q8U757	Q8u757 agrobacteri	763	19	100.0	280	2	Q8FG7	Q8fg7 rhodococcus
691	19	100.0	267	17	Q88H22	Q88h22 pseudomonas	764	19	100.0	280	10	Q9LGD0	Q9lgd0 oryza sativ
692	19	100.0	267	17	Q8ZXX1	Q8zxx1 pyrobaculum	765	19	100.0	280	16	Q9ZLF6	Q9zlf6 rhizobium m
693	19	100.0	268	5	Q8T4T4	Q8t4t4 aedes aegyp	766	19	100.0	280	16	Q8YCH3	Q8ych3 bruceia me
694	19	100.0	268	5	Q8N0R8	Q8n0r8 aedes aegyp	767	19	100.0	280	16	Q8NRY7	Q8nry7 corynebacte
695	19	100.0	268	5	Q9B1G0	Q9b1g0 aedes aegyp	768	19	100.0	280	17	Q8TJ99	Q8tj99 methanosarc
696	19	100.0	268	5	Q8T4T3	Q8t4t3 aedes aegyp	769	19	100.0	281	2	Q9ADV7	Q9adv7 ehrlichia c
697	19	100.0	268	5	Q8T4T3	Q8t4t3 aedes aegyp	770	19	100.0	281	2	Q9ZQ17	Q9zq17 arabidopsis
698	19	100.0	268	5	Q16900	Q16900 aedes aegyp	771	19	100.0	281	10	Q04035	Q04035 arabidopsis
699	19	100.0	268	10	Q8W189	Q8w189 oryza sativ	772	19	100.0	281	12	Q8O1W8	Q8o1w8 squash leaf
700	19	100.0	268	16	Q8BDN4	Q8bdn4 yersinia pe	773	19	100.0	281	16	Q8ZMU7	Q8zmu7 streptomyce
701	19	100.0	269	4	Q8TB51	Q8tb51 homo sapien	774	19	100.0	281	16	Q7UG85	Q7ug85 rhodopirell
702	19	100.0	269	11	Q91VY8	Q91vy8 mus musculu	775	19	100.0	282	2	Q9K3S9	Q9kba9 pseudomonas
703	19	100.0	269	11	Q7TN36	Q7tn36 rattus norv	776	19	100.0	282	2	Q68913	Q68913 streptomyce
704	19	100.0	269	16	Q8FS10	Q8fs10 corynebacte	777	19	100.0	282	16	Q9KGP1	Q9kgp1 bacillus ha
705	19	100.0	269	16	Q7WF26	Q7wf26 bordetella	778	19	100.0	282	16	Q98KW3	Q98kw3 rhizobium l
706	19	100.0	269	16	Q7W3P9	Q7w3p9 bordetella	779	19	100.0	282	16	Q9Z688	Q9z688 clostridium
707	19	100.0	269	16	Q7V8T6	Q7v8t6 bordetella	780	19	100.0	282	16	Q8DZN8	Q8dzn8 streptococ
708	19	100.0	269	17	Q57836	Q57836 pyrococcus	781	19	100.0	282	16	Q89L76	Q89l76 bradyrhizob
709	19	100.0	270	16	Q8N7B3	Q8n7b3 corynebacte	782	19	100.0	282	16	Q8ZDZ5	Q8zdz5 streptomyce
710	19	100.0	271	2	Q86443	Q86443 pseudomonas	783	19	100.0	283	2	Q85983	Q85983 sphingomona
711	19	100.0	271	6	Q9N2J4	Q9nj24 canis faml	784	19	100.0	283	2	Q83VK3	Q83vk3 sphingomona
712	19	100.0	271	8	Q21312	Q21312 tetrahymena	785	19	100.0	284	3	Q13848	Q13848 schizosacch
713	19	100.0	271	16	Q7UNX6	Q7unx6 rhodopirell	786	19	100.0	284	4	Q9HAB8	Q9hab8 homo sapien
714	19	100.0	271	16	Q7UMF0	Q7umf0 rhodopirell	787	19	100.0	284	16	Q8DTC4	Q8dte4 streptococ
715	19	100.0	272	2	Q8G5V5	Q8g5v5 lactobacill	788	19	100.0	284	16	Q8A5T6	Q8ast6 bacteroides
716	19	100.0	272	2	Q8GFH0	Q8gfho rhodococcus	789	19	100.0	285	10	Q9SHM9	Q9shw9 arabidopsis
717	19	100.0	272	5	Q17131	Q17131 babesia equ	790	19	100.0	285	12	Q9PYX4	Q9pyx4 xestia c-ni
718	19	100.0	272	5	Q9T1B0	Q9tyb0 babesia equ	791	19	100.0	285	16	Q8YXL1	Q8yxl1 pseudomonas
719	19	100.0	272	5	Q9TYB1	Q9tyb1 babesia equ	792	19	100.0	286	2	Q05327	Q05327 sphingomona
720	19	100.0	272	5	Q9TYA7	Q9tya7 babesia equ	793	19	100.0	286	13	Q7ZVV6	Q7zvv6 brachydanio
721	19	100.0	272	5	Q9TYB2	Q9tyb2 babesia equ	794	19	100.0	286	16	Q8ZPM8	Q8zpm8 salmonella
722	19	100.0	272	5	Q9TYA9	Q9tya9 babesia equ	795	19	100.0	286	16	Q8ZE21	Q8ze21 yersinia pe
723	19	100.0	272	5	Q9TYA8	Q9tya8 babesia equ	796	19	100.0	286	16	Q8DMF0	Q8dwf0 streptococ
724	19	100.0	272	5	Q46365	Q46365 babesia equ	797	19	100.0	286	16	Q8ZFF1	Q8zff1 pseudomonas
725	19	100.0	272	5	Q9TYB3	Q9tyb3 babesia equ	798	19	100.0	287	5	Q81410	Q81410 caenorhabdi
726	19	100.0	272	5	Q45145	Q45145 caenorhabdi	799	19	100.0	287	16	Q8X649	Q8x649 escherichia
727	19	100.0	273	2	Q9F5V1	Q9f5v1 streptomyce	800	19	100.0	287	16	Q8F719	Q8f719 streptomyce
728	19	100.0	273	5	Q20243	Q20243 caenorhabdi	801	19	100.0	287	16	Q8F8B9	Q8fh89 escherichia
729	19	100.0	273	11	Q8R4X1	Q8r4x1 mus musculu	802	19	100.0	287	16	Q8M34	Q8m34 pseudomonas
730	19	100.0	273	16	Q9FCH8	Q9fch8 streptomyce	803	19	100.0	287	16	Q83KY1	Q83ky1 shigella fl
731	19	100.0	273	17	Q8ZVJ2	Q8zvj2 pyrobaculum	804	19	100.0	288	16	Q8H0U7	Q8hou7 arabidopsis
732	19	100.0	273	17	Q8TXU3	Q8txu3 methanopyru	805	19	100.0	288	16	Q8YCG8	Q8y998 bruceia me
733	19	100.0	274	5	Q20303	Q20303 caenorhabdi	806	19	100.0	288	16	Q81WQ9	Q8lwq9 bacillus an
734	19	100.0	274	11	Q8CDS2	Q8cds2 mus musculu	807	19	100.0	288	16	Q81A22	Q81a22 bacillus ce
735	19	100.0	274	16	Q9KFK4	Q9kfk4 bacillus ha	808	19	100.0	289	11	Q8BMV2	Q8bm2 mus musculu
736	19	100.0	274	16	Q9A112	Q9a112 streptococ	809	19	100.0	289	11	Q8BKP5	Q8bkp5 mus musculu
737	19	100.0	274	16	Q97DB2	Q97db2 clostridium	810	19	100.0	289	16	Q8BKP5	Q8bkp5 mus musculu
738	19	100.0	274	16	Q8P2P0	Q8p2p0 streptococ	811	19	100.0	289	16	Q92MY8	Q92my8 rhizobium m
739	19	100.0	274	16	Q8FTL7	Q8ftl7 corynebacte	812	19	100.0	289	16	Q87IC4	Q87ic4 vibrio para
740	19	100.0	274	16	Q8F513	Q8f513 leptospira	813	19	100.0	290	2	Q84HE4	Q84he4 vibrio vuln
741	19	100.0	274	16	Q879M6	Q879m6 streptococ	814	19	100.0	290	4	Q8TAB2	Q8tab2 homo sapien
742	19	100.0	274	16	Q7VST1	Q7vst1 bordetella	815	19	100.0	290	11	Q9CYZ7	Q9cyz7 mus musculu
743	19	100.0	275	16	Q98H12	Q98h12 rhizobium l	816	19	100.0	290	13	Q8AVB1	Q8avb1 brachydanio
744	19	100.0	275	16	Q97F28	Q97f28 clostridium	817	19	100.0	290	16	Q8UCU3	Q8uc3 agrobacteri
745	19	100.0	275	16	Q82W11	Q82w11 nitrosomona	818	19	100.0	291	2	Q9ZN71	Q9zn71 streptomyce
746	19	100.0	276	5	Q18536	Q18536 caenorhabdi	819	19	100.0	291	12	Q7TLM0	Q7t1m0 choristoneu

820 19 100.0 291 16 Q97LH6 Q97LH6 clostridium  
 821 19 100.0 292 16 Q84633 Q84633 chlamydia t  
 822 19 100.0 293 2 Q33766 Q33766 sphingomona  
 823 19 100.0 294 5 Q16616 Q16616 caenorhabdi  
 824 19 100.0 295 9 Q7Y2C5 Q7Y2C5 bacterioph  
 825 19 100.0 296 5 Q9GNY9 Q9GNY9 leishmania  
 826 19 100.0 297 16 Q83315 Q83315 treponema p  
 827 19 100.0 298 16 Q88LZ3 Q88LZ3 pseudomonas  
 828 19 100.0 299 16 Q826J6 Q826J6 streptomyce  
 829 19 100.0 300 2 Q9RGK4 Q9RGK4 bacteroides  
 830 19 100.0 301 4 Q9NRW3 Q9NRW3 homo sapien  
 831 19 100.0 302 10 Q8SLP1 Q8SLP1 oryza sativ  
 832 19 100.0 303 2 Q9XDK6 Q9XDK6 bacteroides  
 833 19 100.0 304 9 Q854V0 Q854V0 mycobacteri  
 834 19 100.0 305 10 Q94DE7 Q94DE7 oryza sativ  
 835 19 100.0 306 16 Q9PGP2 Q9PGP2 xylella fas  
 836 19 100.0 307 16 Q9KV64 Q9KV64 xylella chol  
 837 19 100.0 308 16 Q8K8N8 Q8K8N8 streptococc  
 838 19 100.0 309 16 Q8AAK1 Q8AAK1 bacteroides  
 839 19 100.0 310 16 Q87ET9 Q87ET9 xylella fas  
 840 19 100.0 311 2 Q9JC87 Q9JC87 nocardioides  
 841 19 100.0 312 17 Q8TU61 Q8TU61 methanosarc  
 842 19 100.0 313 17 Q8PWS7 Q8PWS7 methanosarc  
 843 19 100.0 314 2 Q93LQ8 Q93LQ8 klebsiella  
 844 19 100.0 315 10 Q82163 Q82163 arabisdopsis  
 845 19 100.0 316 16 Q9YV53 Q9YV53 anabaena sp  
 846 19 100.0 317 16 Q8XG56 Q8XG56 anabaena sp  
 847 19 100.0 318 16 Q8XFB0 Q8XFB0 anabaena sp  
 848 19 100.0 319 16 Q8R977 Q8R977 thermoaer  
 849 19 100.0 320 13 Q7ZV21 Q7ZV21 brachydanio  
 850 19 100.0 321 16 Q8U7A1 Q8U7A1 agrobacteri  
 851 19 100.0 322 16 Q8ELP1 Q8ELP1 oceanobacil  
 852 19 100.0 323 16 Q89LM9 Q89LM9 clostridium  
 853 19 100.0 324 2 Q9ADV6 Q9ADV6 ehrlichia c  
 854 19 100.0 325 2 Q9JHS7 Q9JHS7 bradyrhizob  
 855 19 100.0 326 2 Q8GFG6 Q8GFG6 rhodococcus  
 856 19 100.0 327 4 Q9NXA8 Q9NXA8 homo sapien  
 857 19 100.0 328 4 Q9NV14 Q9NV14 homo sapien  
 858 19 100.0 329 4 Q9NQ56 Q9NQ56 homo sapien  
 859 19 100.0 330 4 Q8TC61 Q8TC61 homo sapien  
 860 19 100.0 331 4 Q9NQ48 Q9NQ48 homo sapien  
 861 19 100.0 332 11 Q9JHQS Q9JHQS mus musculu  
 862 19 100.0 333 11 Q8CDG8 Q8CDG8 mus musculu  
 863 19 100.0 334 11 Q8BRX8 Q8BRX8 mus musculu  
 864 19 100.0 335 16 Q9KKW8 Q9KKW8 vibrio chol  
 865 19 100.0 336 16 Q8EIG6 Q8EIG6 shewanella  
 866 19 100.0 337 16 Q8DSV7 Q8DSV7 vibrio vuln  
 867 19 100.0 338 16 Q87G20 Q87G20 vibrio para  
 868 19 100.0 339 17 Q8ZWF0 Q8ZWF0 pyrobaculum  
 869 19 100.0 340 2 Q8ZTV2 Q8ZTV2 pyrobaculum  
 870 19 100.0 341 2 Q8GFG8 Q8GFG8 rhodococcus  
 871 19 100.0 342 16 Q8E3T8 Q8E3T8 streptococc  
 872 19 100.0 343 16 Q8DY72 Q8DY72 streptococc  
 873 19 100.0 344 16 Q7W5F1 Q7W5F1 bordetella  
 874 19 100.0 345 5 Q8LHK7 Q8LHK7 babesia div  
 875 19 100.0 346 10 Q93X14 Q93X14 oryza sativ  
 876 19 100.0 347 16 Q7VW47 Q7VW47 bordetella  
 877 19 100.0 348 17 Q8ZXP5 Q8ZXP5 pyrobaculum  
 878 19 100.0 349 8 Q9SGH8 Q9SGH8 arabisdopsis  
 879 19 100.0 350 10 Q9SGS4 Q9SGS4 arabisdopsis  
 880 19 100.0 351 10 Q8LDJ9 Q8LDJ9 arabisdopsis  
 881 19 100.0 352 16 Q8KAL5 Q8KAL5 chlorobium  
 882 19 100.0 353 5 Q8MZ75 Q8MZ75 drosophila  
 883 19 100.0 354 3 Q81PY4 Q81PY4 drosophila  
 884 19 100.0 355 16 Q8NTU5 Q8NTU5 corynebacte  
 885 19 100.0 356 16 Q7WHF7 Q7WHF7 bordetella  
 886 19 100.0 357 16 Q7W9B0 Q7W9B0 bordetella  
 887 19 100.0 358 16 Q7VYR7 Q7VYR7 bordetella  
 888 19 100.0 359 2 Q9AC88 Q9AC88 pseudomonas  
 889 19 100.0 360 10 Q8LHB8 Q8LHB8 arabisdopsis  
 890 19 100.0 361 11 Q8BN03 Q8BN03 mus musculu  
 891 19 100.0 362 16 Q9PJ91 Q9PJ91 campylobact  
 892 19 100.0 363 2 Q84EL1 Q84EL1 stigmatella

893 19 100.0 306 13 Q7ZX07 Q7ZX07 xenopus lae  
 894 19 100.0 307 16 Q8U746 Q8U746 agrobacteri  
 895 19 100.0 308 16 Q8PGN2 Q8PGN2 xanthomonas  
 896 19 100.0 309 17 Q8ZC24 Q8ZC24 pyrobaculum  
 897 19 100.0 310 5 Q18327 Q18327 caenorhabdi  
 898 19 100.0 311 5 Q8STV9 Q8STV9 encephalito  
 899 19 100.0 312 10 Q9LNR7 Q9LNR7 arabisdopsis  
 900 19 100.0 313 11 Q8VGH1 Q8VGH1 mus musculu  
 901 19 100.0 314 16 Q82PD4 Q82PD4 streptomyce  
 902 19 100.0 315 11 Q8VFG8 Q8VFG8 mus musculu  
 903 19 100.0 316 16 Q9CH02 Q9CH02 lactococcus  
 904 19 100.0 317 16 Q8NL14 Q8NL14 corynebacte  
 905 19 100.0 318 16 Q7V947 Q7V947 prochloroco  
 906 19 100.0 319 10 Q9T000 Q9T000 arabisdopsis  
 907 19 100.0 320 16 Q8VGH8 Q8VGH8 mus musculu  
 908 19 100.0 321 11 Q7TRC3 Q7TRC3 mus musculu  
 909 19 100.0 322 16 Q98RE1 Q98RE1 mycoplasma  
 910 19 100.0 323 17 Q8PYV5 Q8PYV5 methanosarc  
 911 19 100.0 324 4 Q9V6B6 Q9V6B6 homo sapien  
 912 19 100.0 325 10 Q9ZSC1 Q9ZSC1 lycopersico  
 913 19 100.0 326 10 Q9SPD6 Q9SPD6 arabisdopsis  
 914 19 100.0 327 10 Q84Z99 Q84Z99 oryza sativ  
 915 19 100.0 328 11 Q8K2C6 Q8K2C6 mus musculu  
 916 19 100.0 329 12 Q9EMQ9 Q9EMQ9 amsaeta moo  
 917 19 100.0 330 16 Q9CK67 Q9CK67 pasteurella  
 918 19 100.0 331 16 Q88E77 Q88E77 pseudomonas  
 919 19 100.0 332 16 Q87Y83 Q87Y83 pseudomonas  
 920 19 100.0 333 11 Q9FT11 Q9FT11 oryza sativ  
 921 19 100.0 334 11 Q8VF62 Q8VF62 mus musculu  
 922 19 100.0 335 16 Q98N28 Q98N28 rhizobium l  
 923 19 100.0 336 16 Q8NU00 Q8NU00 corynebacte  
 924 19 100.0 337 16 Q8FU71 Q8FU71 corynebacte  
 925 19 100.0 338 11 Q88WP5 Q88WP5 lactobacill  
 926 19 100.0 339 16 Q9HMX6 Q9HMX6 halobacteri  
 927 19 100.0 340 17 Q97ZS5 Q97ZS5 sulfolobus  
 928 19 100.0 341 17 Q8ZT15 Q8ZT15 pyrobaculum  
 929 19 100.0 342 2 Q8GFG9 Q8GFG9 rhodococcus  
 930 19 100.0 343 3 Q12125 Q12125 saccharomyc  
 931 19 100.0 344 16 Q7XTY9 Q7XTY9 oryza sativ  
 932 19 100.0 345 16 Q8NQR2 Q8NQR2 corynebacte  
 933 19 100.0 346 5 Q25791 Q25791 plasmodium  
 934 19 100.0 347 11 Q8VGE0 Q8VGE0 mus musculu  
 935 19 100.0 348 16 Q92X69 Q92X69 rhizobium m  
 936 19 100.0 349 16 Q92AX6 Q92AX6 listeria in  
 937 19 100.0 350 16 Q8Y6K0 Q8Y6K0 listeria mo  
 938 19 100.0 351 16 Q7U8G9 Q7U8G9 synechococc  
 939 19 100.0 352 5 Q8WQX2 Q8WQX2 leishmania  
 940 19 100.0 353 11 Q8VFD7 Q8VFD7 mus musculu  
 941 19 100.0 354 11 Q7TR12 Q7TR12 mus musculu  
 942 19 100.0 355 11 Q7TRY1 Q7TRY1 mus musculu  
 943 19 100.0 356 11 Q7TRU4 Q7TRU4 mus musculu  
 944 19 100.0 357 16 Q8PL19 Q8PL19 xanthomonas  
 945 19 100.0 358 16 Q8P9B2 Q8P9B2 xanthomonas  
 946 19 100.0 359 16 Q8FVU5 Q8FVU5 bruceella su  
 947 19 100.0 360 16 Q88HX8 Q88HX8 pseudomonas  
 948 19 100.0 361 16 Q81J51 Q81J51 bacillus an  
 949 19 100.0 362 16 Q81JG8 Q81JG8 bacillus an  
 950 19 100.0 363 17 Q28600 Q28600 archaeoglob  
 951 19 100.0 364 17 Q8TEB9 Q8TEB9 homo sapien  
 952 19 100.0 365 4 Q8TU97 Q8TU97 pan troglod  
 953 19 100.0 366 6 Q9TSM8 Q9TSM8 macaca fasc  
 954 19 100.0 367 6 Q9TSM7 Q9TSM7 macaca fasc  
 955 19 100.0 368 6 Q9TU88 Q9TU88 gorilla gor  
 956 19 100.0 369 6 Q9TUA0 Q9TUA0 pan troglod  
 957 19 100.0 370 6 Q9TU85 Q9TU85 gorilla gor  
 958 19 100.0 371 10 Q8W5L7 Q8W5L7 oryza sativ  
 959 19 100.0 372 10 Q7XH71 Q7XH71 oryza sativ  
 960 19 100.0 373 11 Q9EQ93 Q9EQ93 mus musculu  
 961 19 100.0 374 5 Q8WS25 Q8WS25 trypanosoma  
 962 19 100.0 375 16 Q9SPE2 Q9SPE2 malus domes  
 963 19 100.0 376 16 Q7UKU5 Q7UKU5 rhodopirell  
 964 19 100.0 377 5 Q9U375 Q9U375 caenorhabdi

966 19 100.0 317 9 Q854X2 mycobacteri  
967 19 100.0 317 10 Q81AG7 arabidopsis  
968 19 100.0 317 11 P97827 rattus norv  
969 19 100.0 317 11 Q92202 cricetus  
970 19 100.0 317 16 Q812Y4 bacillus ce  
971 19 100.0 317 17 Q30168 archaeoglob  
972 19 100.0 318 2 Q9X5F7 zymomonas m  
973 19 100.0 318 2 Q83UD0 neisseria g  
974 19 100.0 318 2 Q7WZ93 nonmuraea  
975 19 100.0 318 3 Q9C135 mortierella  
976 19 100.0 318 10 Q8LCB9 arabidopsis  
977 19 100.0 318 10 Q7XE27 oryza sativ  
978 19 100.0 318 11 Q9QZ21 mus musculu  
979 19 100.0 318 11 Q8VGA9 mus musculu  
980 19 100.0 318 16 Q8FEC0 escherichia  
981 19 100.0 318 16 Q83LQ1 shigella fl  
982 19 100.0 318 17 Q8T7T1 methanosarc  
983 19 100.0 318 17 Q8TQ14 methanosarc  
984 19 100.0 319 3 Q9C0Z8 schizosacch  
985 19 100.0 319 5 Q9STC9 drosophila  
986 19 100.0 319 10 Q9SW32 arabidopsis  
987 19 100.0 319 16 Q8FTS9 corynebacte  
988 19 100.0 319 16 Q83E34 coxiella bu  
989 19 100.0 319 17 Q8HK71 thermoplasm  
990 19 100.0 319 17 Q8TR50 methanosarc  
991 19 100.0 320 10 Q9LD47 arabidopsis  
992 19 100.0 320 10 Q8LFT4 arabidopsis  
993 19 100.0 320 10 Q8LKN3 arabidopsis  
994 19 100.0 320 10 Q9LUF3 arabidopsis  
995 19 100.0 320 10 Q9C972 arabidopsis  
996 19 100.0 320 10 Q84WV9 arabidopsis  
997 19 100.0 320 10 Q9CAH5 arabidopsis  
998 19 100.0 321 2 Q7X1G1 leptospiril  
999 19 100.0 321 4 Q8N6A0 homo sapien  
1000 19 100.0 322 2 Q7WX56 alcaligenes

## ALIGNMENTS

## RESULT 1

Q9TS18 ID Q9TS18 PRELIMINARY; PRT; 20 AA.  
AC Q9TS18; AC Q9TS18; PRELIMINARY; PRT; 30 AA.  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Cytotoxin-binding protein (Fragment).  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94039134; PubMed=7693466;  
RA Lutz F., Mohr M., Grimmig M., Leidolf R., Linder D.;  
RT "Pseudomonas aeruginosa cytotoxin-binding protein in rabbit  
erythrocyte membranes. An oligomer of 28 kDa with similarity to  
transmembrane channel proteins".  
RL Eur. J. Biochem. 217:1123-1128 (1993).  
DR PIR; S39049; S39049.  
DR HSSP; P47865; IJ4N.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005215; F:transporter activity; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR000425; MIP.  
DR Pfam; PF00230; MIP; 1.  
SQ SEQUENCE 20 AA; 2414 MW; 40882D7B5283D2D CRC64;  
Query Match 100.0%; Score 19; DB 6; Length 20;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
RESULT 3  
P82214 ID P82214 PRELIMINARY; PRT; 30 AA.  
AC P82214; AC P82214; PRELIMINARY; PRT; 30 AA.  
DT 01-OCT-2001 (TRENBLrel. 18, Created)  
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Myosin regulatory light chain 2 (MLC-2) (Fragment).  
OS Bombyx mori (Silk moth).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Bombycoidea;  
OC Bombycidae; Bombyx.  
OX NCBI\_TaxID=7091;  
RN [1]  
RP SEQUENCE.  
RX STRAIN=XINHANG X KEMING; TISSUE=Body wall, and Fat body;  
RA Zhong B.X.;  
RT "Protein database for several tissues derived from five instar of  
silkworm".  
RL I Chuan Heuh Pac 28:217-224 (2001).  
CC -!- SUBUNIT: MYOSIN IS AN HEXAMER OF 2 HEAVY CHAINS & 4 LIGHT CHAINS  
(BY SIMILARITY).  
CC -!- MISCELLANEOUS: THIS CHAIN BINDS CALCIUM (BY SIMILARITY).  
CC -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.  
DR GO; GO:0016459; C:myosin; IEA.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0007517; P:muscle development; IEA.  
DR InterPro; IPR002048; EF-hand.  
DR PROSITE; PS00018; EF HAND; PARTIAL.  
KW Myosin; Calcium-binding; Muscle protein.  
FT CA\_BIND 25 >30 BY SIMILARITY.  
FT NON\_TER 30  
SQ SEQUENCE 30 AA; 3369 MW; 5B5287CE9EF538F3 CRC64;  
Query Match 100.0%; Score 19; DB 5; Length 30;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 VAEF 4
13 VAEF 16
13 VAEF 16

RESULT 4
125518
D O25518 PRELIMINARY; PRT; 33 AA.
C O25518
Y 01-JAN-1998 (TREMBlrel. 05, Created)
Y 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
Y 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
Y Hypothetical protein HP0847.
N HP0847.
S Helicobacter pylori (Campylobacter pylori).
C Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
C Helicobacteraceae; Helicobacter.
X NCBI_TaxID=210;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=26695 / ATCC 700392;
X MEDLINE=97394467; PubMed=9252185;
A Tomb J.-F., White O., Kerlavage A.R., Klenk H.-P., Gill S., Dougherty B.A.,
A Fleischmann R.D., Ketchum K.A., Klenk H.-P., Peterson S.,
A Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
A Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
A McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
A Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
A Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
A Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
A Venter J.C.;
T "The complete genome sequence of the gastric pathogen Helicobacter
T pylori."
L Nature 388:539-547(1997).
R EMBL; AE000596; AAD07920.1; -.
R TIGR; G64625; G64625.
W Hypothetical protein; Complete proteome.
Q SEQUENCE 33 AA; 3704 MW; FA5F5631C0DB943 CRC64;

Query Match 100.0%; Score 19; DB 16; Length 33;
Best Local Similarity 100.0%; Pred. No. 5.1e-02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 VAEF 4
15 VAEF 18
15 VAEF 18

RESULT 5
182219
D O82219 PRELIMINARY; PRT; 35 AA.
C O82219
Y 01-MAR-2002 (TREMBlrel. 20, Created)
Y 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
Y 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
Y Hypothetical protein PAE0485.
N PAE0485.
S Pyrobaculum aerophilum.
C Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
X C Thermoproteaceae; Pyrobaculum.
X NCBI_TaxID=13773;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=IM2 / ATCC 51768 / DSM 7523;
X MEDLINE=21664397; PubMed=11792869;
Y Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
Y Miller J.H.;
A "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
A aerophilum."
X Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
```

```
DR EMBL; AE009770; AAL62822.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 35 AA; 3934 MW; A324B4F94F6A14BB CRC64;

Query Match 100.0%; Score 19; DB 17; Length 35;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 VAEF 4
18 VAEF 21
18 VAEF 21

RESULT 6
Q9TQ7
D Q9TQ7 PRELIMINARY; PRT; 36 AA.
C Q9TQ7
Y 01-MAY-2000 (TREMBlrel. 13, Created)
Y 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
Y 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
Y Transferrin (Fragment).
D Transferrin (Fragment).
S Equus caballus (Horse).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
X NCBI_TaxID=9796;
N [1]
P SEQUENCE FROM N.A.
R Ciffard J.M., Brandon R.B., Bell T.K.;
RT "Further identification of single nucleotide polymorphisms in the
RT equine transferrin gene."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF185729; AAF05508.1; -.
DR EMBL; AF185723; AAF05501.1; -.
DR EMBL; AF185724; AAF05503.1; -.
DR EMBL; AF185725; AAF05504.1; -.
DR EMBL; AF185726; AAF05505.1; -.
DR EMBL; AF185727; AAF05506.1; -.
DR EMBL; AF185728; AAF05507.1; -.
DR HSSP; P19134; 1TFD.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008199; F:ferric iron binding; IEA.
DR GO; GO:0006879; P:iron ion homeostasis; IEA.
DR GO; GO:0006826; P:iron ion transport; IEA.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 1.
FT NON_TER 1 36
FT NON_TER 36
SQ SEQUENCE 36 AA; 3802 MW; E2DD122186D5FA8 CRC64;

Query Match 100.0%; Score 19; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 VAEF 4
28 VAEF 31
28 VAEF 31

RESULT 7
Q48420
D Q48420 PRELIMINARY; PRT; 44 AA.
C Q48420
Y 01-NOV-1996 (TREMBlrel. 01, Created)
Y 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
Y 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
Y Hypothetical protein (Fragment).
C Klebsiella pneumoniae.
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Klebsiella.
X NCBI_TaxID=573;
N [1]
P SEQUENCE FROM N.A.
```

```

RC STRAIN=ATCC 25955;
RA Willard B.L.;
RT "Investigation of the Klebsiella pneumoniae 1,3-propanediol pathway:
RT Characterization and expression of glycerol dehydratase and 1,3-
RT propanediol oxidoreductase.";
RL Thesis (1994), Chemical Engineering, University of Wisconsin-Madison.
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=ATCC 25955;
RA Skraly F.A., Willard B.L., Cameron D.C.;
RT "The dha regulon of Klebsiella pneumoniae.";
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).
DR EMBL; U30903; AAT74254.1; -.
DR HSSP; P11244; 1FX8.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000425; MIP.
DR Pfam; PF00230; MIP; 1.
DR ProDom; PD000295; MIP family; 1.
KW Hypothetical protein; Transmembrane; Transport.
FT NON_TER 44 44
FT SEQUENCE 44 AA; 4537 MW; 39851658FF88E734 CRC64;
SQ
Query Match 100.0%; Score 19; DB 2; Length 44;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAEF 4
DB 12 VAEF 15
RESULT 8
Q830H9 PRELIMINARY; PRT; 45 AA.
AC Q830H9;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN EF2805.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seehadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Unayam L., Brinkac L., Seanan M.,
RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074 (2003).
DR EMBL; AB016955; AAC82500.1; -.
DR TIGR; EF2805; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 45 AA; 5278 MW; 89CB55F5CB4C014A CRC64;
Query Match 100.0%; Score 19; DB 16; Length 45;
Best Local Similarity 100.0%; Pred. No. 6.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAEF 4
DB 21 VAEF 24

```

```

RESULT 9
Q25478 PRELIMINARY; PRT; 48 AA.
AC Q25478;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hypothetical protein HP0789.
GN HP0789.
OS Helicobacter pylori (Campylobacter pylori)
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., B.A.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547 (1997).
DR EMBL; AE000591; AAD07847.1; -.
DR PIR; E84618; E64618.
DR TIGR; HP0789; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 48 AA; 5492 MW; ED5911D96F57BFEB CRC64;
Query Match 100.0%; Score 19; DB 16; Length 48;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAEF 4
DB 15 VAEF 18
RESULT 10
Q70XL5 PRELIMINARY; PRT; 51 AA.
AC Q70XL5;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RB1255.
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL; EX294135; CAD7191.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 51 AA; 5951 MW; 60F62389C0FDA6A7 CRC64;
Query Match 100.0%; Score 19; DB 16; Length 51;

```

Best Local Similarity 100.0%; Pred. No. 7.8e+02; Mismatches 0; Gaps 0;

Matches 4; Conservative 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
|  
|  
|  
|  
31 VAEF 34

RESULT 11

ID Q44941 PRELIMINARY; PRT; 52 AA.

AC Q44941; (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DE Hypothetical protein.

GN Short ORF.

LN LON.

OS *Bacillus brevis* (*Brevibacillus brevis*).

OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; *Brevibacillus*.

OX NCBI\_TaxID=1393;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HED31;

RX Ito K., Uda K., Yamagata H.;

RA "Cloning, characterization, and inactivation of the *Bacillus brevis*

RA lon gene.";

RA J. Bacteriol. 174:2281-2287(1982).

RA ENBL; D00863; BAA00736.1; -.

RA PR; A42375; A42375.

RQ SEQUENCE 52 AA; 5558 MW; 5D689099F84F9AAF CRC64;

Query Match 100.0%; Score 19; DB 2; Length 52;

Best Local Similarity 100.0%; Pred. No. 7.9e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
|  
|  
|  
|  
47 VAEF 50

RESULT 12

ID Q8X3Y8 PRELIMINARY; PRT; 54 AA.

AC Q8X3Y8; (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE Hypothetical protein z2382.

GN *Escherichia coli* O157:H7.

OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; *Escherichia*.

OX NCBI\_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grobbeck E.J., Davis N.W., Lim A., Dinalanta E.T., Potamouzis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RA "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";

RA Nature 409:529-533(2001).

RA ENBL; AE005369; AAG56425.1; -.

RA PR; E85745; E85745.

RQ SEQUENCE 54 AA; 6507 MW; 8BFACD88E6140CE CRC64;

Query Match 100.0%; Score 19; DB 16; Length 54;

Best Local Similarity 100.0%; Pred. No. 8.2e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
|  
|  
|  
|  
32 VAEF 35

RESULT 13

ID Q8PIF3 PRELIMINARY; PRT; 54 AA.

AC Q8PIF3; (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Hypothetical protein.

GN C1469.

OS *Escherichia coli* O6.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; *Escherichia*.

OX NCBI\_TaxID=217992;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O6:H1 / CFT073 / ATCC 700928;

RX MEDLINE=22388234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,

RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,

RA Mayhew G.F., Rose D.J., Zhou S., Schwartz F.R.;

RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;

RA "Extensive mosaic structure revealed by the complete genome sequence

RA of uropathogenic *Escherichia coli*.";

RA Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).

RA ENBL; AB016759; AAN79938.1; -.

RQ SEQUENCE 54 AA; 5322 MW; 6B3CD0F21ED4376 CRC64;

Query Match 100.0%; Score 19; DB 16; Length 54;

Best Local Similarity 100.0%; Pred. No. 8.2e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
|  
|  
|  
|  
4 VAEF 7

RESULT 14

ID Q8VSE8 PRELIMINARY; PRT; 61 AA.

AC Q8VSE8; (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Hypothetical protein.

GN CP0186.

OS *Shigella flexneri* 2a.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; *Shigella*.

OX NCBI\_TaxID=42897;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=301;

RA Jin Q., Zhang J.Y., Liu H., Yang J., Yang F., Zhang X.B., Wang J.H.,

RA Yang G.W., Wu H.T., Dong J., Sun L.L., Xue Y., Zhao A.L., Gao Y.S.,

RA Zhu J.P., Kan B., Chen S.X., Yao Z.J., He B.K., Chen R.S., Ma D.L.,

RA Yuan Z.H., Xu J.G., Wang Y., Shen Y., Lu W.C., Qiang B.Q., Wen Y.M.,

RA Hou Y.D.;

RA "Complete DNA sequence and analysis of the large virulence plasmid

RA pCP301 of *Shigella flexneri*.";

RA Submitted (MAY-2001) to the ENBL/GenBank/DBJ databases.

RA ENBL; AF386526; AAL72547.1; -.

RA GO; GO:0046821; C:extrachromosomal DNA; IEA.

RQ SEQUENCE 61 AA; 6558 MW; F1CC17B10B28CBFC CRC64;

Query Match 100.0%; Score 19; DB 2; Length 61;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+02; Indels 0; Gaps 0;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB 54 VAEF 57

RESULT 15

Q9P166 PRELIMINARY; PRT; 61 AA.

AC Q9P166;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DE PRO2435  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Zhang C., Yu Y., Zhang S., Wei H., Zhang Y., Zhou G., Bi J., Liu M.,  
 RA He F.;  
 RT "Functional prediction of the coding sequences of 79 new genes deduced  
 RT by analysis of cDNA clones from human fetal liver.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF119881; AAF69635.1; -;  
 SQ SEQUENCE 61 AA; 7467 MW; 64339504657C1D4 CRC64;

Query Match 100.0%; Score 19; DB 4; Length 61;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB 30 VAEF 33

RESULT 16

Q99GZ9 PRELIMINARY; PRT; 68 AA.

AC Q99GZ9;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DE ORF49 (Hypothetical protein) (Unknown).  
 OS Helicoverpa zea single nucleocapsid nucleopolyhedrovirus,  
 OS Helicoverpa armigera nucleopolyhedrovirus G4, and  
 OS Helicoverpa armigera nuclear polyhedrosis virus.  
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 OC Nucleopolyhedrovirus.  
 OX NCBI\_TaxID=10468, 148363, 51313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa zea single nucleocapsid nucleopolyhedrovirus;  
 RA Chen X., Zhang W.-J., Wong J., Chun G., Lu A., McCutchen B.F.,  
 RA Fresnail J.K., Herrmann R., Dolan M., Tingey S., Hu Z.-H., Vlask J.M.;  
 RT "Genome sequence analysis of Helicoverpa zea single nucleocapsid  
 RT nucleopolyhedrovirus.";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;  
 RA Deng F., Chen X., Vlask J.M., Arif B.M., Hu Z.;  
 RT "Sequence analysis of the GP37 gene of Heliothis armigera single-  
 RT nucleocapsid nucleopolyhedrovirus.";  
 RL Zhongguo Bingduxue 15:35-42 (2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;

RA Wang H., Hu Z., Sun X., Vlask J.M., Chen X.;  
 RT "Sequence analysis of the jap3 gene of Heliothis armigera single-  
 RT nucleocapsid nucleopolyhedrovirus.";  
 RL Zhongguo Bingduxue 15:43-49 (2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;  
 RX MEDLINE=21078302; PubMed=11210934;  
 RA Wang H., Chen X., Wang H., Arif B.M., Vlask J.M., Hu Z.;  
 RT "Nucleotide sequence and transcriptional analysis of a putative basic  
 RT DNA-binding protein of Helicoverpa armigera polyhedrovirus.";  
 RL Virus Genes 22:113-120 (2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;  
 RX MEDLINE=21064569; PubMed=11125177;  
 RA Chen X., IJkel W.F., Tarchini R., Sun X., Sandbrink H., Wang H.,  
 RA Peters S., Zuidema D., Lankhorst R.K., Vlask J.M., Hu Z.;  
 RT "The sequence of the Helicoverpa armigera single-nucleocapsid  
 RT nucleopolyhedrovirus genome.";  
 RL J. Gen. Virol. 82:241-257 (2001).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;  
 RA Chen X., IJkel W.F., Tarchini R., Sun X., Sandbrink H., Wang H.,  
 RA Peters S., Zuidema D., Lankhorst R.K., Vlask J.M., Hu Z.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nuclear polyhedrosis virus; STRAIN=C1;  
 RX PubMed=12050807;  
 RA Zhang C.X., Wu J.C.;  
 RT "Genome structure and the p10 gene of the Helicoverpa armigera  
 RT nucleopolyhedrovirus.";  
 RL Sheng Wu Hua Hsueh Yu Sheng Wu Wu Li Hsueh Pao 33:179-184 (2001).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nuclear polyhedrosis virus; STRAIN=C1;  
 RA Zhang C.X., Jin W.R.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nuclear polyhedrosis virus;  
 RA Fang M., Hu Z., Chen X., Vlask J.M.;  
 RT "Genetic organization of the HindIII-L region of Helicoverpa armigera  
 RT single-nucleocapsid nucleopolyhedrovirus.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF334030; AAL56194.1; -;  
 DR EMBL: AF271059; AAG53791.1; -;  
 DR EMBL: AF303045; AAK96298.1; -;  
 DR EMBL: AF266694; AAK64316.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 68 AA; 7962 MW; 61B7718FBFB195FF CRC64;

Query Match 100.0%; Score 19; DB 12; Length 68;  
 Best Local Similarity 100.0%; Pred. No. 1e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB 29 VAEF 32

RESULT 17

O82X12 PRELIMINARY; PRT; 68 AA.

AC O82X12;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Helix-turn-helix protein, CopG family.  
 DE NE0289.  
 GN Nitrosomonas europaea.

```

JC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
XC Nitrosomonadaceae; Nitrosomonas.
XX NCBI_TaxID=915;
XN [1]
XP SEQUENCE FROM N.A.
XQ MEDLINE=22586410; PubMed=12700255;
A Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
A Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
A Acierio D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
T "Complete genome sequence of the ammonia-oxidizing bacterium and
T obligate chemolithoautotroph Nitrosomonas europaea.";
L J. Bacteriol. 185:2759-2773(2003).
R EMBL; BX321857; CAD84200.1; -.
W Complete proteome.
XQ SEQUENCE 68 AA; 7800 MW; 8A939B537524DD6 CRC64;

Query Match 100.0%; Score 19; DB 16; Length 69;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4
b 31 VAEF 34

RESULT 18
7X7L1 PRELIMINARY; PRT; 69 AA.
D Q7X7L1
C Q7X7L1
T 01-OCT-2003 (TrEMBLrel. 25, Created)
T 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
T OSUNBA0044M19.22 protein (OSUNBA0053B21.1 protein).
N OSUNBA0044M19.22 OR OSUNBA0053B21.1.
S Oryza sativa (Rice).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
C Ehrhartoideae; Oryzaceae; Oryza.
X NCBI_TaxID=4530;
XN [1]
XP SEQUENCE FROM N.A.
A Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
A Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
A Liu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
A Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
A Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
A Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
A Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
A Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
A Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
A Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
R EMBL; AL731601; CAB05035.1; -.
R EMBL; AL731599; CAB05527.1; -.
XQ SEQUENCE 69 AA; 7267 MW; 77839333BC873F0BE CRC64;

Query Match 100.0%; Score 19; DB 10; Length 69;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4
b 31 VAEF 34

RESULT 19
28NPA7 PRELIMINARY; PRT; 69 AA.
D Q8NPA7
C Q8NPA7;
T 01-OCT-2002 (TrEMBLrel. 22, Created)
T 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
T 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
T Hypothetical protein Cgll1907.

```

```

GN CGL1907.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
XN [1]
XP SEQUENCE FROM N.A.
XQ STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
A Nakagawa S.;
R "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
R Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
R EMBL; AF005280; BAB99300.1; -.
R Hypothetical protein; Complete proteome.
XQ SEQUENCE 69 AA; 7815 MW; C0E2A072C2295DD2 CRC64;

Query Match 100.0%; Score 19; DB 16; Length 69;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4
b 23 VAEF 26

RESULT 20
Q8AUQ2 PRELIMINARY; PRT; 71 AA.
D Q8AUQ2
C Q8AUQ2;
T 01-MAR-2003 (TrEMBLrel. 23, Created)
T 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
T Transferrin (Fragment).
S Salmo trutta (Brown trout).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Actinopterygii; Neopterygii; Teleostei; Euteleostei;
C Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
X NCBI_TaxID=8032;
XN [1]
XP SEQUENCE FROM N.A.
R STRAIN=Str-1, Str-2, Str-3, Str-4, Str-9, Str-10, and Str-11;
R MEDLINE=22135992; PubMed=12140239;
R Antunes A., Templeton A.R., Guyomard R., Alexandrino P.;
R "The role of nuclear genes in intraspecific evolutionary inference:
R genealogy of the transferrin gene in the brown trout.";
R Mol. Biol. Evol. 19:1272-1287(2002).
R EMBL; AF488850; AAN17027.1; -.
R EMBL; AF488849; AAN17027.1; JOINED.
R EMBL; AF488858; AAN17032.1; -.
R EMBL; AF488857; AAN17032.1; JOINED.
R EMBL; AF488865; AAN17037.1; -.
R EMBL; AF488865; AAN17037.1; JOINED.
R EMBL; AF488874; AAN17042.1; -.
R EMBL; AF488873; AAN17042.1; JOINED.
R EMBL; AF488913; AAN17068.1; -.
R EMBL; AF488913; AAN17068.1; JOINED.
R EMBL; AF488922; AAN17073.1; -.
R EMBL; AF488921; AAN17073.1; JOINED.
R EMBL; AF488930; AAN17078.1; -.
R EMBL; AF488929; AAN17078.1; JOINED.
R GO; GO:0005576; C:extracellular; IEA.
R GO; GO:0008199; F:ferric iron binding; IEA.
R GO; GO:0006879; P:iron ion homeostasis; IEA.
R GO; GO:0006826; P:iron ion transport; IEA.
R InterPro; IPR001156; Transferrin.
R Pfam; PF00405; Transferrin; 1.
R PRINTS; PR00422; TRANSFERRIN.
R NON_TER 1
R NON_TER 71
R SEQUENCE 71 AA; 7546 MW; 892E38F2D2DD2AC4D CRC64;

Query Match 100.0%; Score 19; DB 13; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 VAEF 4
Db      33 VAEF 36

RESULT 21
Q8AUPO  PRELIMINARY; PRT; 71 AA.
AC Q8AUPO;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE Transferrin (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ssa-1, and Ssa-2;
RX MEDLINE=22135992; PubMed=12140239;
RA Atunes A., Templeton A.R., Guyomard R., Alexandrino P.;
RT "The role of nuclear genes in intraspecific evolutionary inference:
RL genealogy of the transferrin gene in the brown trout.";
DR ENBL; AF488834; AAN17017.1; -.
DR ENBL; AF488833; AAN17017.1; JOINED.
DR ENBL; AF488842; AAN17022.1; -.
DR ENBL; AF488841; AAN17022.1; JOINED.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008199; F:ferrous iron binding; IEA.
DR GO; GO:0006879; P:piron ion homeostasis; IEA.
DR GO; GO:0006826; P:piron ion transport; IEA.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; Transferrin; 1.
DR PRINTS; PR00422; TRANSFERRIN.
FT NON TER 1
FT NON TER 71
SQ SEQUENCE 71 AA; 7504 MW; 9D86A1234CBAC4D CRC64;

Query Match 100.0%; Score 19; DB 13; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 VAEF 4
Db      33 VAEF 36

RESULT 22
P97252  PRELIMINARY; PRT; 72 AA.
AC P97252;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DE Late control gene D protein (Fragment).
OS D.
OC Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KL2;
RX MEDLINE=97251359; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampaio G., Seki Y.,
RA Sivasubraman S., Tagami H., Takeda J., Takemoto K., Wada C.,

```

```

RA Yamamoto Y., Horiuchi T.;
RT "A 460-Kb DNA sequence of the Escherichia coli K-12 genome
RL corresponding to the 40.1-50.0 min region on the linkage map.";
RNA Res. 3:379-392(1996).
DR ENBL; D90847; BAA15949.1; -.
DR ENBL; D90846; BAA15938.1; -.
FT NON TER 1
SQ SEQUENCE 72 AA; 8099 MW; 42A63B25B00EADCE CRC64;

Query Match 100.0%; Score 19; DB 2; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 VAEF 4
Db      3 VAEF 6

RESULT 23
Q97JU4  PRELIMINARY; PRT; 74 AA.
AC Q97JU4;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein CAC1179.
GN CAC1179.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 782 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Bretton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR ENBL; AB007632; AAK79151.1; -.
DR PIR; D97045; D97045.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 74 AA; 8747 MW; 633633CB0A0C293A CRC64;

Query Match 100.0%; Score 19; DB 16; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 VAEF 4
Db      5 VAEF 8

RESULT 24
Q82MX0  PRELIMINARY; PRT; 74 AA.
AC Q82MX0;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein.
GN SAV1532.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;

```

A Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
A Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
A Kikuchi H., Shiba T., Sakaki Y., Hattori M.,  
T "Genome sequence of an industrial microorganism Streptomyces  
T avermitilis: deducing the ability of producing secondary  
T metabolites";  
L Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
N [2]  
P SEQUENCE FROM N.A.  
C STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
X MEDLINE=22608306; PubMed=12692562;  
A Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
A Sakaki Y., Hattori M., Omura S.,  
T "Complete genome sequence and comparative analysis of the industrial  
T microorganism Streptomyces avermitilis";  
L Nat. Biotechnol. 21:526-531(2003).  
R EMBL; AP005027; BAC69243.1; --  
W Hypothetical protein; Complete proteome.  
Q SEQUENCE 74 AA; 7960 MW; D8CE71C808AFB01 CRC64;  
Query Match 100.0%; Score 19; DB 16; Length 74;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 VAEF 4  
b 36 VAEF 39  
RESULT 25  
1862M8  
D Q862M8 PRELIMINARY; PRT; 77 AA.  
AC Q862M8;  
YC 01-JUN-2003 (TReMBLrel. 24, Created)  
YT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
VT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Similar to MAD2 protein (Fragment).  
IS Bos taurus (Bovine).  
XC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
XC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
XC Bovidae; Bovinae; Bos.  
XN NCBI\_TaxID=9913;  
IN [1]  
IP SEQUENCE FROM N.A.  
IX MEDLINE=22544902; PubMed=12658628;  
JA Ishiwata H., Katsuma S., Kizaki K., Patel O.V., Nakano H.,  
JA Takahashi T., Inai K., Hirasawa A., Shiojima S., Ikawa H., Suzuki Y.,  
JA Tsujimoto G., Izaike Y., Todoroki J., Hashizume K.,  
T "Characterization of gene expression profiles in early bovine  
T pregnancy using a custom cDNA microarray";  
L Mol. Reprod. Dev. 65:9-18(2003).  
R EMBL; AB098952; BAC56442.1; --  
R InterPro; IPR003511; DNABind\_HORVA.  
R PROSITE; PS00815; HORVA; 1.  
T NON TER 1  
T NON TER 77  
Q SEQUENCE 77 AA; 8832 MW; E2E1D625C848E571 CRC64;  
Query Match 100.0%; Score 19; DB 6; Length 77;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 VAEF 4  
b 6 VAEF 9  
RESULT 26  
188VS7  
D Q88VS7 PRELIMINARY; PRT; 77 AA.  
AC Q88VS7;  
YT 01-JUN-2003 (TReMBLrel. 24, Created)  
VT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)

DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN LP 1960.  
OS Lactobacillus plantarum.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
L.L. NCBI\_TaxID=1590;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=NCIMB 8826 / WCFSL;  
RX MEDLINE=22480296; PubMed=12566566;  
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,  
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,  
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,  
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,  
RA De Vos W.M., Siezen R.J.,  
RT "Complete genome sequence of Lactobacillus plantarum WCFS1";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).  
RW EMBL; AL935257; CAD64344.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 77 AA; 8618 MW; C00E2BF1D401F2F CRC64;  
Query Match 100.0%; Score 19; DB 16; Length 77;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VAEF 4  
Db 68 VAEF 71  
RESULT 27  
Q98582  
ID Q98582 PRELIMINARY; PRT; 79 AA.  
AC Q98582;  
DT 01-FEB-1997 (TReMBLrel. 02, Created)  
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
DE A532L protein.  
GN A532L.  
OS Paramyxium bursaria chlorella virus 1 (PBCV-1).  
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.  
OX NCBI\_TaxID=10506;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=96400190; PubMed=8806566;  
RA Kutish G.F., Li Y., Lu Z., Furuta M., Rock D.L., Van Etten J.L.;  
RT "Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map  
RT positions 182 to 258";  
RL Virology 223:303-317(1996).  
RN [2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=20013326; PubMed=1054099;  
RA Kaiser A., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,  
RA Lisec A.D., Nickerson K.W., Van Etten J.L.;  
RT "Chlorella virus PBCV-1 encodes a functional homosperridine  
RT synthase";  
RL Virology 263:254-262(1999).  
RN [3]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=20478054; PubMed=11021991;  
RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;  
RT "Characterization of a beta-1,3-glucanase encoded by chlorella virus  
RT PBCV-1";  
RL Virology 276:27-36(2000).  
RN [4]  
RN SEQUENCE FROM N.A.  
RA Van Etten J.L.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RN SEQUENCE FROM N.A.  
RA Van Etten J.L.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.

```

RN [6]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBSJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBSJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBSJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RA Gurnon J.R., Graves M.V., Van Etten J.L.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; U42580; AAC96899.1; -.
DR PIR; T18034; T18034.
SQ SEQUENCE 79 AA; 8698 MW; B191C627F5D5C5A7 CRC64;

Query Match 100.0%; Score 19; DB 12; Length 79;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 49 VAEF 52

RESULT 28
Q9KI36 PRELIMINARY; PRT; 80 AA.
AC Q9KI36;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Hypothetical protein BH2182.
DE Ysg.
GN Ysg.
OS Agrobacterium tumefaciens.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=358;
RN [1]
RP SEQUENCE FROM N.A.
RA Winans S.C., Zhu J., Oger P.M., Schrammeijer B., Hooykaas P.J.,
RA Farrand S.K.;
RT "Octopine-type Ti plasmid sequence."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF242881; AAF77180.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Plasmid.
SQ SEQUENCE 80 AA; 8723 MW; 4E8379C1D907CA8E CRC64;

Query Match 100.0%; Score 19; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 65 VAEF 68

RESULT 29
Q89WL6 PRELIMINARY; PRT; 81 AA.
AC Q89WL6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Bsr0662 protein.
GN Bsr0662.
OS Bradyrhizobium japonicum.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpō S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005937; BAC45927.1; -.
KW Complete proteome.
SQ SEQUENCE 81 AA; 9235 MW; D512A9FFED0DA6C7 CRC64;

Query Match 100.0%; Score 19; DB 16; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 43 VAEF 46

RESULT 30
Q9KAV4 PRELIMINARY; PRT; 84 AA.
AC Q9KAV4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein BH2182.
GN BH2182.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Maeni N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001514; BAB05901.1; -.
DR PIR; F83922; F83922.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 84 AA; 9669 MW; 6CDE3768ED9FD84 CRC64;

Query Match 100.0%; Score 19; DB 16; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 38 VAEF 41

RESULT 31
Q821D2 PRELIMINARY; PRT; 84 AA.
AC Q821D2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Hypothetical protein.
GN CCA01010.
OS Chlamydomophila caviae.
```

C Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.  
 X NCBI\_TaxID=83557;  
 N [1]  
 P SEQUENCE FROM N.A.  
 C STRAIN=GPIC;  
 X MEDLINE=22569155; PubMed=12682364;  
 A Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,  
 A Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A.,  
 A Unayam L.A., Haft D.H., Peterson J., Beaman M.J., White O.,  
 A Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavov P.M.,  
 A Fraser C.M.,  
 T "genome sequence of Chlamydophila caviae (Chlamydia psittaci GPIC):  
 T examining the role of niche-specific genes in the evolution of the  
 T Chlamydiaceae.";  
 L Nucleic Acids Res. 31:2134-2147(2003).  
 R EMBL; AB016997; AAP05749.1; --  
 R TIGR; CCA01010; --  
 W Hypothetical protein; Complete proteome.  
 Q SEQUENCE 84 AA; 9191 MW; 7DF6009729C7093A CRC64;  
  
 Query Match 100.0%; Score 19; DB 16; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 Y 1 VAEF 4  
 b 68 VAEF 71  
  
 RESULT 32  
 7VPV6  
 D Q7VPV6 PRELIMINARY; PRT; 84 AA.  
 C Q7VPV6;  
 T 01-OCT-2003 (TrEMBLrel. 25, Created)  
 T 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 E Hypothetical protein.  
 N CFA0775.  
 S Chlamydia pneumoniae (Chlamydophila pneumoniae).  
 C Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.  
 X NCBI\_TaxID=83558;  
 N [1]  
 P SEQUENCE FROM N.A.  
 C STRAIN-TW-183;  
 A Geng M.M., Schumacher A., Muehlendorfer I., Bensch K.W., Schaefer K.P.,  
 A Schneider S., Pohl T., Essig A., Marre R., Melchers K.;  
 T "The genome sequence of Chlamydia pneumoniae TW183 and comparison with  
 T other Chlamydia strains based on whole genome sequence analysis.";  
 T Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 R EMBL; AB017159; AAP98704.1; --  
 W Hypothetical protein.  
 Q SEQUENCE 84 AA; 9260 MW; 7786DF8BDD7B3AB8 CRC64;  
  
 Query Match 100.0%; Score 19; DB 16; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 Y 1 VAEF 4  
 b 68 VAEF 71  
  
 RESULT 33  
 043116  
 D 043116 PRELIMINARY; PRT; 85 AA.  
 C 043116;  
 T 01-JUN-1998 (TrEMBLrel. 06, Created)  
 T 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 E Hypothetical protein (fragment).  
 S Colletotrichum gloeosporioides (Anthraxnose fungus) (Glomerella  
 S singulata).  
 C Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;  
 OC Glomerella.  
 OX NCBI\_TaxID=5457;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 C STRAIN=VQ62;  
 RA Stephenson S.-A., Maclean D.J., Manners J.M.;  
 RT "Disruption of a novel pathogenicity gene of Colletotrichum  
 RT gloeosporioides results in a hypersensitive response in the host  
 RT Stylosanthes guianensis.";  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U94183; A892222.1; --  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR001387; HTH\_3.  
 DR Pfam; PF01381; HTH\_3; 1.  
 DR SMART; SM00530; HTH\_XRE; 1.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 85 AA; 9316 MW; BFB9A0E5F44E9CF2 CRC64;  
  
 Query Match 100.0%; Score 19; DB 3; Length 85;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 VAEF 4  
 Db 45 VAEF 48  
  
 RESULT 34  
 Q99149  
 ID Q99149 PRELIMINARY; PRT; 87 AA.  
 AC Q99149;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Transferrin (fragments).  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 C TISSUE=LIVER;  
 RX MEDLINE=91177867; PubMed=1848850;  
 RA Tu G.F., Achen M.G., Aldred A.K., Southwell B.R., Schreiber G.;  
 RT "The distribution of cerebral expression of the transferrin gene is  
 RT species specific.";  
 RL J. Biol. Chem. 266:6201-6208(1991).  
 CC -I- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH  
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING  
 CC OF AN ANION, USUALLY BICARBONATE.  
 DR EMBL; M64691; AAA31585.1; --  
 DR EMBL; M64692; AAA31586.1; --  
 DR PIR; A38725; A38725.  
 DR HSSP; P19134; 1TFD.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0008199; F:ferric iron binding; IEA.  
 DR GO; GO:0006879; P:iron ion homeostasis; IEA.  
 DR GO; GO:0008262; P:iron ion transport; IEA.  
 DR InterPro; IPR001156; Transferrin.  
 DR SMART; SM00094; TR\_FER; 1.  
 DR PROSITE; PS00205; TRANSFERRIN\_1;  
 KW Iron transport; Metal-binding.  
 FT NON\_TER 1  
 FT NON\_CONS 44 45  
 FT NON\_TER 87 87  
 SQ SEQUENCE 87 AA; 9433 MW; AA464B2ABDC92FAB CRC64;  
  
 Query Match 100.0%; Score 19; DB 6; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 VAEF 4
Db 13 VAEF 16

RESULT 35
Q7Y3Y9 PRELIMINARY; PRT; 87 AA.
ID Q7Y3Y9
AC Q7Y3Y9;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Bacteriophage PY54.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=172667;
RN [1]
RP SEQUENCE FROM N.A.
RA Hertwig S., Klein I., Schmidt V., Beck S., Hammerl J.A., Appel B.;
RT "Sequence analysis of the genome of the temperate Yersinia
RT enterocolitica phage PY54."
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hertwig S.;
RT Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ564013; CAD91786.1; -.
KW Exonuclease; Hypothetical protein.
SQ SEQUENCE 87 AA; 8538 MW; D17D7A7E3075459B CRC64;
Query Match 100.0%; Score 19; DB 9; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
Db 31 VAEF 34

RESULT 36
Q9S590 PRELIMINARY; PRT; 88 AA.
ID Q9S590
AC Q9S590;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE Cytochrome P460 (Fragment).
CN CYP.
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
RN [1]
RP SEQUENCE FROM N.A.
RA Iizumi T., Nakamura K.;
RT "Construction of tryptophan requiring-mutant of Nitrosomonas europaea
RT by inactivation of trpC gene using homologous recombination."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB030031; BAA83388.1; -.
PT NON TER 88
SQ SEQUENCE 88 AA; 9765 MW; A59003348F4E280C CRC64;
Query Match 100.0%; Score 19; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
Db 29 VAEF 32

RESULT 37
Q41185 PRELIMINARY; PRT; 88 AA.
ID Q41185;
AC Q41185;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Glyceraldehyde-3-phosphate-dehydrogenase subunit GapB (Fragment).
GN GAPB.
OS Arabidopsis thaliana (Mouse-ear cress).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93013005; PubMed=1398114;
RA Shih M.C., Heinrich P., Goodman H.M.;
RT "Cloning and chromosomal mapping of nuclear genes encoding chloroplast
RT and cytosolic glyceraldehyde-3-phosphate-dehydrogenase from
RT Arabidopsis thaliana";
RL Gene 119:317-319(1992).
DR EMBL; S45911; AAB29353.1; -.
DR GO; GO:0009507; C:chloroplast; IEA.
KW Chloroplast.
FT NON TER 88
SQ SEQUENCE 88 AA; 9198 MW; FCDB6A7E37B6999D CRC64;
Query Match 100.0%; Score 19; DB 10; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
Db 38 VAEF 41

RESULT 38
Q9ZGY2 PRELIMINARY; PRT; 89 AA.
ID Q9ZGY2
AC Q9ZGY2;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN Y1103.
OS Yersinia pestis.
OG Plasmid pMT-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM10+;
RX MEDLINE=99043898; PubMed=9826348;
RA Lindler L.E., Plano G.V., Burland V., Mayhew G.F., Blattner F.R.;
RT "Complete DNA sequence and detailed analysis of the Yersinia pestis
RT KIM5 plasmid encoding murine toxin and capsular antigen."
RL Infect. Immun. 66:5731-5742(1998).
DR EMBL; AF074611; AAC82761.1; -.
DR PIR; T15018; T15018.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR003006; IG_MHC.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 89 AA; 10736 MW; 72B4221126023EE4 CRC64;
Query Match 100.0%; Score 19; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4

```

b 3 VAEF 6  
|||||

## RESULT 39

9CND1 PRELIMINARY; PRT; 89 AA.  
D Q9CNX1  
C Q9CNX1  
T 01-JUN-2001 (TrEMBLrel. 17, Created)  
I 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
I 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
E RPS15.  
N RPS15 OR PM0301.  
S Pasteurella multocida.  
N Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
C Pasteurellaceae; Pasteurella.  
C NCBI\_TaxID=747;  
X N [1]  
P SEQUENCE FROM N.A.  
F STRAIN=Pm70;  
C MEDLINE=21145866; PubMed=11248100;  
K May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
I "Complete genomic sequence of Pasteurella multocida Pm70.";  
L Proc. Natl Acad. Sci. U.S.A. 98:3460-3465 (2001).  
R ENBL; AS006066; AAK02385.1; -.  
R HSP; P80378; IAB3.  
R GO; GO:0005622; C:intracellular; IEA.  
R GO; GO:0005840; C:ribosome; IEA.  
R GO; GO:0003735; P:structural constituent of ribosome; IEA.  
R GO; GO:0006412; P:protein biosynthesis; IEA.  
R InterPro; IPR000589; Ribosomal\_S15.  
R InterPro; IPR005290; Ribosomal\_S15\_b.  
R Pfam; PF00312; Ribosomal\_S15; 1.  
R ProDom; PD157043; RS15\_bact; 1.  
R TIGRFAMs; TIGR00952; S15\_bact; 1.  
R PROSITE; PS00362; RIBOSOMAL\_S15; 1.  
M Complete proteome.  
Q SEQUENCE 89 AA; 10184 MW; F79620AB5263ED0E CRC64;

Query Match 100.0%; Score 19; DB 16; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
|||||  
b 12 VAEF 15

## RESULT 40

9CFD1 PRELIMINARY; PRT; 89 AA.  
D Q3CFD1  
C Q3CFD1  
T 01-JUN-2001 (TrEMBLrel. 17, Created)  
I 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
I 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
E Hypothetical protein ypfB.  
N YPFB OR IL1550.  
S Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
N Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
X NCBI\_TaxID=1360;  
N [1]  
P SEQUENCE FROM N.A.  
C STRAIN=IL1403;  
MEDLINE=21235186; PubMed=11337471;  
A Bolotin A., Wincker P., Mauer S., Jallion O., Malarne K.,  
A Weissenbach J., Ehrlich S.D., Sorokin A.;  
T "The complete genome sequence of the lactic acid bacterium Lactococcus  
T lactis ssp. lactis IL1403.";  
L Genome Res. 11:731-753 (2001).  
L ENBL; AE006385; AAK05648.1; -.  
R PIR; F86818; F86818.  
R Hypothetical protein; Complete proteome.  
Q SEQUENCE 89 AA; 10403 MW; D90DF3C71D12D06 CRC64;

Query Match 100.0%; Score 19; DB 16; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||||  
Db 68 VAEF 71

## RESULT 41

Q81ZX3 PRELIMINARY; PRT; 90 AA.  
ID Q81ZX3  
AC Q81ZX3  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE MAD2 mitotic arrest deficient-like 1 variant.  
GN MAD2L1  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Vin F., Fan D.M.;  
RT Identifying a new variant of MAD2L1.";  
RL Submitted (JUN-2001) to the ENBL/GenBank/DBJ databases.  
DR EMBL; AF394735; AAN74648.1; -.  
DR InterPro; IPR003511; DNABind\_HORMA.  
DR Pfam; PF02301; HORMA; 1.  
DR PROSITE; PS50815; HORMA; 1.  
SQ SEQUENCE 90 AA; 10335 MW; 8209F5A7A7D8D09B CRC64;

Query Match 100.0%; Score 19; DB 4; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||||  
Db 20 VAEF 23

## RESULT 42

Q98TB2 PRELIMINARY; PRT; 91 AA.  
ID Q98TB2  
AC Q98TB2  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Preproinsulin (Fragment).  
OS Ambloplites rupestris (Rock bass).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Centrarchidae; Acanthopterygii; Percomorpha; Perciformes; Percoidae;  
OC NCBI\_TaxID=109273;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA al-Mahrouki A.A., Irwin D.M., Youson J.H.;  
RT "Molecular cloning of preproinsulin cDNA from the rock bass.";  
RL Submitted (OCT-1999) to the ENBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
DR ENBL; AF199584; AAK28708.1; -.  
DR HSP; P01308; ILPH.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0005179; F:hormone activity; IEA.  
DR GO; GO:0007582; P:physiological processes; IEA.  
DR InterPro; IPR004825; Ins/IGF/relax.  
DR Pfam; PF00049; Insulin; 1.  
DR PRINTS; PR00277; INSULIN.  
DR SMART; SM00078; ILGF; 1.

DR PROSITE; PS00262; INSULIN; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 91 91  
 SQ SEQUENCE 91 AA; 10100 MW; E86C8B256DC69D39 CRC64;

Query Match 100.0%; Score 19; DB 13; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB 55 VAEF 58

## RESULT 43

Q8KY12 PRELIMINARY; PRT; 94 AA.

AC Q8KY12; 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Putative transcriptional regulator (Fragment).  
 OS Rhizobium celi.  
 OG Plasmid p42b.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
 OX NCBI\_TaxID=29449;  
 RN [1]\_TaxID=29449;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CE3;  
 RA Cevallos M.A.; Izquierdo J.; Porta H.; Tun C.; Davila G.; Brom S.;  
 RT "Rhizobium etli CE3 contains at least three plasmids of the RepABC  
 family: A structural and an evolutionary analysis."  
 RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF313446; RAM89940.1; -;  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.

KW Plasmid.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 94 AA; 10222 MW; C09D6EA3DFF1E2BA CRC64;

Query Match 100.0%; Score 19; DB 2; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB 38 VAEF 41

## RESULT 44

O46425 PRELIMINARY; PRT; 94 AA.

AC O46425;  
 DT 01-JUN-1998 (TRENBLrel. 06, Created)  
 DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Aquaporin 1 (Fragment).  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand;  
 RA Carter E.P.; Umenishi F.; Matthay M.A.; Verkman A.S.;  
 RT "Increased water permeability across the blood-gas barrier in rabbit  
 lungs in the first 24 hours after birth."  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).  
 DR EMBL; AF000311; AAE94408.1; -;  
 DR HSSP; P25972; IFQY;  
 DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0005741; C:mitochondrial outer membrane; IEA.  
 DR GO; GO:0015288; F:porin activity; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR000425; MIP.  
 DR Pfam; PF00230; MIP; 1.  
 DR ProDom; PD000295; MIP family; 1.  
 DR PROSITE; PS00221; MIP; 1.  
 KW Porin; Transmembrane; Transport.  
 FT NON\_TER 1 1  
 FT NON\_TER 94 94  
 SQ SEQUENCE 94 AA; 10057 MW; 963D5527631B8CDC CRC64;

Query Match 100.0%; Score 19; DB 6; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB 9 VAEF 12

## RESULT 45

Q8U8Q1 PRELIMINARY; PRT; 96 AA.

AC Q8U8Q1;  
 DT 01-JUN-2002 (TRENBLrel. 21, Created)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Hypothetical protein Atu4039.  
 GN ATU4039.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
 OX NCBI\_TaxID=176299;  
 RN [1]\_TaxID=176299;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W.; Setubal J.C.; Kaul R.; Monks D.E.; Kitajima J.P.; Woo L.;  
 RA Okura Y.K.; Zhou Y.; Chen L.; Wood G.E.; Almeida N.F. Jr.; Woo L.;  
 RA Chen Y.; Paulsen I.T.; Eisen J.A.; Karp P.D.; Bovee D. Sr.;  
 RA Chapman P.; Clendenning J.; Deatherage G.; Gillet W.; Grant C.;  
 RA Kutayavin T.; Levy R.; Li M.-J.; McClelland E.; Palmieri A.;  
 RA Raymond C.; Rouse G.; Saenphimmachak C.; Wu Z.; Romero P.; Gordon D.;  
 RA Zhang S.; Yoo H.; Tao Y.; Biddle P.; Jung M.; Krespan W.; Perry M.;  
 RA Gordon-Kamm B.; Liao L.; Kim S.; Hendrick C.; Zhao Z.-Y.; Dolan M.;  
 RA Chumley F.; Tingey S.V.; Tomb J.-F.; Gordon M.F.; Olson M.V.;  
 RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 C58."  
 RL Science 294:2317-2323(2001).  
 DR EMBL; AE009334; AAL44840.1; -;  
 DR PIR; AE3053; AB3053.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 96 AA; 11193 MW; FBD635894B46A8AE CRC64;

Query Match 100.0%; Score 19; DB 16; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 DB 55 VAEF 58

## RESULT 46

Q89WX7 PRELIMINARY; PRT; 96 AA.

ID Q89WX7;  
 AC Q89WX7;  
 DT 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Bsl0551 protein.

N BSL0551.  
 S Bradyrhizobium japonicum.  
 C Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 C Bradyrhizobiaceae; Bradyrhizobium.  
 X NCBI\_TaxID=375;  
 N N [1]  
 P SEQUENCE FROM N.A.  
 C STRAIN=USDA 110;  
 X MEDLINE=22484998; PubMed=12597275;  
 A Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
 A Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,  
 A Kohara M., Matsumoto M., Shampo S., Tsuruoka H., Wada T., Yamada M.,  
 A Tabata S.;  
 T "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 T Bradyrhizobium japonicum USDA110.";  
 L DNA Res. 9:189-197(2002).  
 R EMBL; AP005936; C:membrane; IEA.  
 R GO; GO:0016020; C:membrane; IEA.  
 R InterPro; IPR003425; Unk\_YGGT.  
 R Pfam; PF02325; YGGT; 1.  
 W Complete proteome.  
 Q SEQUENCE 96 AA; 10976 MW; A9E11F9AA4BC7734 CRC64;

Query Match 100.0%; Score 19; DB 16; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 b 44 VAEF 47  
 |||||  
 RESULT 47  
 8U0G9 PRELIMINARY; PRT; 96 AA.  
 C Q8U0G9  
 T 01-JUN-2002 (TrEMBLrel. 21, Created)  
 T 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 T 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 E Hypothetical protein PF1620.

S Pyrococcus furiosus.  
 C Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 C Pyrococcus.  
 X NCBI\_TaxID=2261;  
 N N [1]  
 P SEQUENCE FROM N.A.  
 C STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;  
 A Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
 T "The complete sequence of the Pyrococcus furiosus genome.";  
 L Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 R EMBL; AE010462; AAL81744.1; -.  
 W Hypothetical protein; Complete proteome.  
 Q SEQUENCE 96 AA; 10804 MW; D6DAE09D096D577A CRC64;

Query Match 100.0%; Score 19; DB 17; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 b 24 VAEF 27  
 |||||  
 RESULT 48  
 8G6535 PRELIMINARY; PRT; 97 AA.  
 C Q8G6535  
 T 01-NOV-1996 (TrEMBLrel. 01, Created)  
 T 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 T 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 E 2C/3A (Fragment).  
 N 2C/3A.

OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LSH/S;  
 RX MEDLINE=92348853; PubMed=1668326;  
 RA Fineschi N., Cavallieri F., Garelick H., Prugnot A., Pellegrini V.,  
 RA Zuckerman A.J.;  
 RT "Characterization of a hepatitis A virus strain suitable for vaccine  
 RT production.";  
 RL J. Hepatol. 13:S146-S151(1991).  
 DR EMBL; S44109; AAB22740.2; -.  
 FT NON TER 1 1  
 FT NON TER 97 97  
 SQ SEQUENCE 97 AA; 10614 MW; D207F7CB93110DDF CRC64;

Query Match 100.0%; Score 19; DB 12; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 Db 44 VAEF 47  
 |||||  
 RESULT 49  
 Q8Y476 PRELIMINARY; PRT; 97 AA.  
 ID Q8Y476  
 AC Q8Y476  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein lmo2579.

OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1639;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EGD-e / Serovar 1/2a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Donann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Entian K.-D., Fslhi H., Garcia-del Portillo P., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunat F., Kurapkat G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 RA Varquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
 RT "Comparative Genomics of Listeria species.";  
 RL Science 294:849-852(2001).  
 DR EMBL; AL591983; CAD00657.1; -.  
 DR PIR; AC1397; AC1397.  
 DR ListiList; LMO02579; -.  
 DR InterPro; IPR007138; ABM.  
 DR Pfam; PF03992; ABM; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 97 AA; 10979 MW; 9E758586E94218E0 CRC64;

Query Match 100.0%; Score 19; DB 16; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 Db 15 VAEF 18  
 |||||

RESULT 50  
Q9HZ38 ID Q9HZ38 PRELIMINARY; PRT; 99 AA.  
AC Q9HZ38; (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Hypothetical protein PA3202.  
GN PA3202.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.B.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen."  
RL Nature 406:959-964(2000).  
DR EMBL; AE004744; AAG06590.1; --  
DR PIR; E83244; E83244.  
DR InterPro; IPR005545; YCII.  
DR Pfam; PF03795; YCII; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 99 AA; 10609 MW; 3AD945F4D54485C CRC64;  
Query Match 100.0%; Score 19; DB 16; Length 99;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
DB 62 VAEF 65  
RESULT 51  
Q985M3 ID Q985M3 PRELIMINARY; PRT; 99 AA.  
AC Q985M3; (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Conserved hypothetical protein.  
GN PSPT01808.  
OS Pseudomonas syringae (pv. tomato).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=323;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DC3000;  
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,  
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,  
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,  
RA Brinko L., Beanan M., Haft D., Selengut J., Nelson W., Davidson T.,  
RA White O., Fraser C., Collier A.;  
RT "Complete sequence of Pseudomonas syringae";  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB016863; AAO55328.1; --  
DR TIGR; PSPT01808; --  
DR InterPro; IPR005545; YCII.  
DR Pfam; PF03795; YCII; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 99 AA; 10504 MW; 62836007E4849392 CRC64;  
Query Match 100.0%; Score 19; DB 16; Length 99;

Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
DB 62 VAEF 65  
RESULT 52  
Q82UR2 ID Q82UR2 PRELIMINARY; PRT; 100 AA.  
AC Q82UR2; (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN NEI419.  
OS Nitrosomonas europaea.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;  
OC Nitrosomonadaceae; Nitrosomonas.  
OX NCBI\_TaxID=915;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19718 / IFO 14298;  
RX MEDLINE=22585410; PubMed=12700255;  
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,  
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,  
RA Arclero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;  
RT "Complete genome sequence of the ammonia-oxidizing bacterium and  
RT obligate chemolithoautotroph Nitrosomonas europaea";  
RL J. Bacteriol. 185:2759-2773(2003).  
DR EMBL; BX321861; CAD85330.1; --  
DR InterPro; IPR005545; YCII.  
DR Pfam; PF03795; YCII; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 100 AA; 10836 MW; 82DC1153B8FAB27 CRC64;  
Query Match 100.0%; Score 19; DB 16; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
DB 62 VAEF 65  
RESULT 53  
Q8VS54 ID Q8VS54 PRELIMINARY; PRT; 101 AA.  
AC Q8VS54; (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Elongation factor TS (Fragment).  
GN TSF.  
OS Lactobacillus reuteri.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=1598;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 55739;  
RA Nam S.J., Kim J.K., Park J.Y., Ha Y.L., Kim J.H.;  
RT "Cloning of UMP-kinase gene from Lactobacillus reuteri ATCC 55739";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF401482; AAL60142.1; --  
DR GO; GO:0003746; P:translation elongation factor activity; IEA.  
DR GO; GO:0006414; P:translational elongation; IEA.  
DR InterPro; IPR001816; EF TS.  
DR Pfam; PF00889; EF TS; 1.  
FT NON TER 1  
SQ SEQUENCE 101 AA; 11650 MW; BBCF30D941DB2B9B CRC64;

Query Match 100.0%; Score 19; DB 2; Length 101;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
62 VAEF 65

RESULT 54  
Q7YC2 PRELIMINARY; PRT; 101 AA.  
Q7YC2; 2001 (TREMBlrel. 18, Created)  
01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
01-OCT-2003 (TREMBlrel. 24, Last annotation update)  
Hypothetical protein SSO1404.  
N SSO1404.  
S Sulfolobus solfataricus.  
C Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
C Sulfolobus.  
X X NCBI\_TaxID=2287;  
N [1]  
P SEQUENCE FROM N.A.  
C STRAIN-ATCC 35092 / DSM 1617 / P2;  
MEDLINE=21332296; PubMed=11427726;  
A She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
A Wayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,  
A De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,  
A Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
A Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
A Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
A Garret P.A., Ragan M.A., Sengen C.W., Van der Oost J.,  
A "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."  
L Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
R EMBL; AE006755; AA41639.1; -;  
R PIR; H90297; H90297.  
R InterPro; IPR003799; DUF196.  
R Pfam; PF02647; DUF196; 1.  
R TIGRFAMs; TIGR01573; cas2; 1.  
Q Hypothetical protein; Complete proteome.  
Q SEQUENCE 101 AA; 11936 MW; A7C338AD76202E17 CRC64;

Query Match 100.0%; Score 19; DB 17; Length 101;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
20 VAEF 23

RESULT 55  
Q81R27 PRELIMINARY; PRT; 102 AA.  
Q81R27;  
01-JUN-2003 (TREMBlrel. 24, Created)  
01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
Hypothetical protein.  
N BA2232.  
S Bacillus anthracis (strain Ames).  
C Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
X X NCBI\_TaxID=198094;  
N [1]  
P SEQUENCE FROM N.A.  
R MEDLINE=22608414; PubMed=12721629;  
A Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,  
A Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
A Holtzapple E.K., Ostad O.A., Helgason E., Ristone J., Wu M.,  
A Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,  
A DeBoy R.T., Madupu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
A Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,

RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,  
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolisto A.-B.,  
RA Fraser C.M.;  
RT "The genome sequence of Bacillus anthracis Ames and comparison to  
RT closely related bacteria."  
RL Nature 423:81-86(2003).  
DR EMBL; AE017031; AAP26109.1; -;  
DR TIGR; BAZ232; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 102 AA; 11925 MW; 13B8E296C8A341B1 CRC64;

Query Match 100.0%; Score 19; DB 16; Length 102;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
54 VAEF 57

RESULT 56  
Q7Z2N2 PRELIMINARY; PRT; 103 AA.  
Q7Z2N2;  
01-OCT-2003 (TREMBlrel. 25, Created)  
01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
BK150C2.6 (Putative novel protein similar to APOBEC1 (Apolipoprotein B  
BK150C2.6 (Putative novel protein similar to APOBEC1 (Apolipoprotein B  
BK150C2.6 (Putative novel protein similar to APOBEC1 (Apolipoprotein B  
BK150C2.6 (Putative novel protein similar to APOBEC1 (Apolipoprotein B  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Matthews L.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DDJ databases.  
DR EMBL; AL022318; CAB45276.1; -;  
KW Lipoprotein.  
FT NON\_TER 1 103  
SQ SEQUENCE 103 AA; 12146 MW; 5DC969AE3ED348D9 CRC64;

Query Match 100.0%; Score 19; DB 4; Length 103;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
46 VAEF 49

RESULT 57  
Q9F368 PRELIMINARY; PRT; 103 AA.  
Q9F368;  
01-MAR-2001 (TREMBlrel. 16, Created)  
01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
Hypothetical protein SCO4420.  
GN SCO4420 OR SC6F11.18.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
R STRAIN-A3(2) / M145;  
RC MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Coble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S., Taylor K.,  
 RA Seeger K., Saunders D., Sharp S., Squares K., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,  
 RA "Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939120; CAC38429.1; --  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 103 AA; 11387 MW; 55C2B8059EB75B5 CRC64;

Query Match 100.0%; Score 19; DB 16; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 Db 54 VAEF 57

RESULT 58  
 Q89MS9 PRELIMINARY; PRT; 103 AA.  
 ID Q89MS9;  
 AC Q89MS9;  
 DT 01-JUN-2003 (TReMBLrel. 24, Created)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
 DE Hypothetical protein; Complete proteome.  
 DE BLR4113 protein.  
 EN BLR4113.  
 CS Bradyrhizobium japonicum.  
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 CC Bradyrhizobiaceae; Bradyrhizobium.  
 CX NCBI\_TaxID=375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=USDA 110;  
 RX MEDLINE=22484998; PubMed=12597275;  
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
 RA Sasamoto S., Watanabe A., Ideawa K., Iriguchi M., Kawashima K.,  
 RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,  
 RA Tabata S.;  
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 RT Bradyrhizobium japonicum USDA110.";  
 RL DNA Res. 9:189-197(2002).  
 DR EMBL; AP005950; BAC49378.1; --  
 KW Complete proteome.  
 SQ SEQUENCE 103 AA; 11104 MW; B227CB9F90FB723 CRC64;

Query Match 100.0%; Score 19; DB 16; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 Db 96 VAEF 99

RESULT 59  
 Q856B5 PRELIMINARY; PRT; 104 AA.  
 ID Q856B5;  
 AC Q856B5;  
 DT 01-JUN-2003 (TReMBLrel. 24, Created)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 DE Gp57.  
 CS Mycobacteriophage Barnyard.  
 CC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
 CX NCBI\_TaxID=205880;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=22592660; PubMed=12705866;  
 RA Pedulla M.L., Ford M.E., Hout J.M., Karthikeyan T., Wadsworth C.,  
 RA Lewis J.A., Jacobs-Sera D., Falbo J., Gross J., Pannunzio N.R.,  
 RA Brucker W., Kumar V., Kandasamy J., Keenan L., Bardarov S.,  
 RA Kravoy J., Lawrence J.G., Jacobs W.R. Jr., Hendrix R.W.,  
 RA Hatfull G.F.;  
 RT "Origins of highly mosaic mycobacteriophage genomes.";  
 RL Cell 113:171-182(2003).  
 DR EMBL; AY129339; AAN02111.1; --  
 SQ SEQUENCE 104 AA; 11074 MW; BB806EAC401E9FD0 CRC64;

Query Match 100.0%; Score 19; DB 9; Length 104;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 Db 59 VAEF 62

RESULT 60  
 Q8SVK0 PRELIMINARY; PRT; 105 AA.  
 ID Q8SVK0;  
 AC Q8SVK0;  
 DT 01-JUN-2002 (TReMBLrel. 21, Created)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Hypothetical protein ECU05\_0770.  
 GN ECU05\_0770.  
 OS Encephalitozoon cuniculi.  
 CC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.  
 CX NCBI\_TaxID=6035;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RA Genoscope;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RX MEDLINE=21576510; PubMed=11719806;  
 RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,  
 RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,  
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,  
 RA Weissenbach J., Vivares C.P.;  
 RT "Genome sequence and gene compaction of the eukaryote parasite  
 RT Encephalitozoon cuniculi.";  
 RL Nature 414:450-453(2001).  
 DR EMBL; AL590445; CAD26596.1; --  
 KW Hypothetical protein.  
 SQ SEQUENCE 105 AA; 11897 MW; 08059108C05D3BB9 CRC64;

Query Match 100.0%; Score 19; DB 5; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 Db 77 VAEF 80

RESULT 61  
 Q856U1 PRELIMINARY; PRT; 105 AA.  
 ID Q856U1;  
 AC Q856U1;  
 DT 01-JUN-2003 (TReMBLrel. 24, Created)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 DE Gp3.  
 CS Mycobacteriophage Corndog.  
 CC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
 CX NCBI\_TaxID=205875;  
 RN [1]

N P SEQUENCE FROM N.A.  
 X MEDLINE=22592660; PubMed=12705866;  
 A Pedulla M.L., Ford M.E., Houtz J.M., Karthikeyan T., Wadsworth C.,  
 A Lewis J.A., Jacobs-Sera D., Falbo J., Gross J., Panunzio N.R.,  
 A Brucker W., Kumar V., Kandasamy J., Keenan L., Bardarov S.,  
 A Kriakov J., Lawrence J.G., Jacobs W.R. Jr., Hendrix R.W.,  
 A Hatfull G.F.;  
 T "Origins of highly mosaic mycobacteriophage genomes.";  
 L Cell 113:171-182(2003).  
 R ENBL; AY129335; AA01935.1; -. 7D0EC03F5F2AF2DD CRC64;  
 Q SEQUENCE 105 AA; 11955 MW; 7D0EC03F5F2AF2DD CRC64;  
  
 Query Match 100.0%; Score 19; DB 9; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 Y 1 VAEF 4  
 b 41 VAEF 44  
  
 RESULT 62  
 D Q96Z31 PRELIMINARY; PRT; 105 AA.  
 C Q96Z31;  
 T 01-DEC-2001 (TrEMBLrel. 19, Created)  
 T 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 E Hypothetical protein ST2000.  
 N ST2000.  
 S Sulfolobus tokodaii.  
 C Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 C Sulfolobus.  
 X NCBI\_TaxID=111955;  
 N [1]  
 P SEQUENCE FROM N.A.  
 C STRAIN=JCM 10545 / 7;  
 X MEDLINE=21456156; PubMed=11572479;  
 A Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
 A Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,  
 A Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 A Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,  
 A Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
 A Ohnita T., Kikuchi H.;  
 T "Complete genome sequence of an aerobic thermoacidophilic  
 Crenarchaeon, Sulfolobus tokodaii strain7.";  
 L DNA Res. 8:123-140(2001).  
 R ENBL; AP000988; BAB67095.1; -.  
 R InterPro; IPR006783; Transposase\_29.  
 R Pfam; PF04693; Transposase\_29; 2-  
 W Hypothetical protein; Complete proteome.  
 Q SEQUENCE 105 AA; 11997 MW; AFE6ABD4DF2DEC0E CRC64;  
  
 Query Match 100.0%; Score 19; DB 17; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 Y 1 VAEF 4  
 b 29 VAEF 32  
  
 RESULT 63  
 D P71530 PRELIMINARY; PRT; 106 AA.  
 C P71530;  
 T 01-FEB-1997 (TrEMBLrel. 02, Created)  
 T 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
 T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 E NifX.  
 N NifX.  
 S Methanococcus maripaludis.

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanococcaceae; Methanococcus.  
 X NCBI\_TaxID=39152;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LL;  
 RX MEDLINE=96011360; PubMed=7592322;  
 RA Blank C.E., Kessler P.S., Leigh J.A.;  
 RT "Genetics in methanogens: transposon insertion mutagenesis of a  
 RT Methanococcus maripaludis nifH gene.";  
 RL J. Bacteriol. 177:5773-5777(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LL;  
 RX MEDLINE=97144542; PubMed=8990309;  
 RA Kessler P.S., McLarnan J., Leigh J.A.;  
 RT "Nitrogenase phylogeny and the molybdenum dependence of nitrogen  
 RT fixation in Methanococcus maripaludis.";  
 RL J. Bacteriol. 179:541-543(1997).  
 DR ENBL; U75887; AAC45519.1; -.  
 DR PIR; T10097; T10097.  
 DR InterPro; IPR003731; DUF153.  
 DR Pfam; PF02579; Nitro\_FeMo-Co; 1.  
 SQ SEQUENCE 106 AA; 11862 MW; 4734D89CFED33F3B CRC64;  
  
 Query Match 100.0%; Score 19; DB 1; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 VAEF 4  
 Db 35 VAEF 38  
  
 RESULT 64  
 Q9BGL5 PRELIMINARY; PRT; 106 AA.  
 ID Q9BGL5;  
 AC Q9BGL5;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DE Matrixin (Fragment).  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OC NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Endometrium;  
 RA Smith G.W., Ricke W.A., Cassar C.A., Smith M.F.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR ENBL; AF267158; AAG59846.1; -.  
 DR HSP; P09237; LMQ.  
 DR MEROPS; M10.008; -.  
 DR GO; GO:0005578; C:extracellular matrix; IEA.  
 DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR006026; Peptidase M.  
 DR InterPro; IPR001818; Pept\_M10A\_M12B.  
 DR Pfam; PF0413; Peptidase\_M10; 1.  
 DR PRINTS; PR00138; Peptidase\_M10; 1.  
 DR SMART; SM00235; ZmNC; 1.  
 DR NON\_TER 1  
 FT NON TER 106  
 SQ SEQUENCE 106 AA; 11578 MW; EBD271054928018 CRC64;  
  
 Query Match 100.0%; Score 19; DB 6; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 VAEF 4  
 Db 35 VAEF 38

```

Db          6 VAEF 9
Query Match 100.0%; Score 19; DB 3; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 65
ID O23682 PRELIMINARY; PRT; 106 AA.
AC O23682;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Cathepsin B-like cysteine proteinase (Fragment).
GN T7123.13...
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Federspiel N.A., Conway A.B., Conway A.R., Davis K., Brendel V.,
RA Palm C.J., Au M., Araujo R., Chung E., Kurtz D.B., Buehler E.,
RA Dewar K., Peng J., Kim C., Li Y., Shimm P., Sun H., Oji O.,
RA Osborne B., Shen Y.K., Toriumi M., Vyctskaia V., Yu G., Theologis A.,
RA Zcker J., Davis R.W.;
RT "Genomic sequence of Arabidopsis BAC T7123."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBSJ databases.
DR EMBL; U99595; AAC24377.1; -.
FT NON TER 106 106
SQ SEQUENCE 106 AA; 11743 MW; C5D5EA5897AD17AF CRC64;

Query Match 100.0%; Score 19; DB 10; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 75 VAEF 78

RESULT 66
Q9C1L1 PRELIMINARY; PRT; 107 AA.
ID Q9C1L1;
AC Q9C1L1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative thioredoxin G6G8.7.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=211335672; PubMed=11238395;
RA Bean L.E., Dvorachek W.H. Jr., Braun E.L., Errett A., Saenz G.S.,
RA Giles M.D., Werner-Washburne M., Nelson M.A., Natvig D.O.;
RT "Analysis of the pdx-1 (enz-1/sno-1) region of the Neurospora crassa
genome. Correlation of pyridoxine-requiring phenotypes with mutations
in two structural genes."
RL Genetics 157:1067-1075 (2001).
DR EMBL; AF309689; AAK07845.1; -.
DR HSP; P80028; ITOF.
DR GO; GO:0005489; P:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR006662; Thiorid.
DR InterPro; IPR006663; Thioridox_dom2.
DR Pfam; PF00085; Thiorid; 1.
DR PRINTS; PR00421; THIOREDOXIN.
DR Redox-active center.
KW Redox-active center.
SQ SEQUENCE 107 AA; 11676 MW; CF45EAF85BE3776 CRC64;

Query Match 100.0%; Score 19; DB 5; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 40 VAEF 43

RESULT 67
Q8IHV3 PRELIMINARY; PRT; 107 AA.
ID Q8IHV3;
AC Q8IHV3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF11_0423.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S.L., Craig A., Kyes S.,
RA Eissen J.A., Rutherford K., Salzberg S.L., Suh B., Peterson J., Angiuoli S.,
RA Chant M., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perle M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum."
RL Nature 419:498-511 (2002).
DR EMBL; AE014842; AAN36006.1; -.
KW Hypothetical protein.
SQ SEQUENCE 107 AA; 12714 MW; CA351258FCF46EF5 CRC64;

Query Match 100.0%; Score 19; DB 5; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 40 VAEF 43

RESULT 68
Q8NRX6 PRELIMINARY; PRT; 107 AA.
ID Q8NRX6;
AC Q8NRX6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Uncharacterized ACR.
GN CGL0915.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF005276; BAB96308.1; -.
DR InterPro; IPR007138; ABM.
DR Pfam; PF03992; ABM; 1.
KW Complete proteome.

```

IQ SEQUENCE 107 AA; 12373 MW; B5ACF23621078018 CRC64;  
 Query Match 100.0%; Score 19; DB 16; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 1 VAEF 4  
 b 22 VAEF 25

RESULT 69  
 9PU82 PRELIMINARY; PRT; 109 AA.  
 C Q9PU82  
 T 01-MAY-2000 (TrEMBLrel. 13, Created)  
 T 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 E Ratol binding protein (fragment).  
 S Crocodylus niloticus (Nile crocodile).  
 C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 C Archosauria; Crocodylia; Crocodylidae; Crocodylinae; Crocodylus.  
 X NCBI\_TaxID=8501;  
 N [1]  
 P SEQUENCE FROM N.A.  
 C TISSUE=Liver;  
 X MEDLINE=20022983; PubMed=10555283;  
 X Hughes S., Zelus D., Mouchiroud D.;  
 T "Warm-blooded isochore structure in Nile crocodile and turtle.";  
 L Mol. Biol. Evol. 18:1521-1527(1999).  
 R EMBL; AJ011392; CAB56418.1; -;  
 R HSP; P02753; IRBP.  
 R GO; GO:0005215; P:transporter activity; IEA.  
 R GO; GO:0006810; P:transport; IEA.  
 R InterPro; IPR002345; Lipocalin.  
 R InterPro; IPR000566; Lipocalin\_cyfPAP.  
 R Pfam; PF00061; lipocalin; 1.  
 R PRINTS; PR00179; LIPOCALIN.  
 T NON TER 1  
 T NON TER 109 109  
 Q SEQUENCE 109 AA; 12552 MW; ODCAE93895B91A23 CRC64;  
 Query Match 100.0%; Score 19; DB 13; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 1 VAEF 4  
 b 25 VAEF 28

RESULT 70  
 53333 PRELIMINARY; PRT; 109 AA.  
 C O53333  
 T 01-JUN-1998 (TrEMBLrel. 06, Created)  
 T 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 E Hypothetical protein RV3183.  
 N RV3183 OR MTV014.27 OR MT3275.  
 S Mycobacterium tuberculosis.  
 C Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 C Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 X NCBI\_TaxID=1773;  
 N [1]  
 P SEQUENCE FROM N.A.  
 C STRAIN=H37RV;  
 X MEDLINE=98295987; PubMed=9634230;  
 X Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 X Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,  
 X Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 X Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 X Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL021646; CAAL648.1; -;  
 DR EMBL; AE007140; AAK47615.1; -;  
 DR PIR; E70949; E70949.  
 DR TIGR; MT3275; -;  
 DR TubercuList; RV3183; -;  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR001387; HTH\_3.  
 DR Pfam; PF01381; HTH\_3; 1.  
 DR SMART; SMO0530; HTH\_XRE; 1.  
 DR KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 109 AA; 11802 MW; A66DBE160D79613A CRC64;  
 Query Match 100.0%; Score 19; DB 16; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 DB 97 VAEF 100

RESULT 71  
 Q7TX28 PRELIMINARY; PRT; 109 AA.  
 AC Q7TX28;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Possible transcriptional regulatory protein.  
 GN MB3209.  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AF2122/97;  
 RX MEDLINE=22709107; PubMed=12788972;  
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,  
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,  
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RT "The complete genome sequence of Mycobacterium bovis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
 DR EMBL; BX248345; CAD95301.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 109 AA; 11802 MW; A66DBE160D79613A CRC64;

Query Match 100.0%; Score 19; DB 16; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VAEF 4  
 DB 97 VAEF 100

RESULT 72  
ID O11343 PRELIMINARY; PRT; 110 AA.  
AC O11343;  
DT 01-JUL-1997 (TREMBLrel. 04, Created)  
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE XI-5 protein (Fragment).  
GN xi-5.  
OS Molluscum contagiosum virus subtype 1 (MCV1).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Molluscipoxvirus.  
OX NCBI\_TaxID=10280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Moratilla M., Agromayor M., Nunez A., Funes J.M., Varas A.J.,  
RA Lopez-Esteban J.L., Esteban M., Martin-Gallardo A.,  
RT "A Random DNA Sequencing, Computer-Based Approach for the Generation  
RT of a Gene Map of Molluscum Contagiosum Virus.",  
RL Virus Genes 0:0-0(1997).  
DR EMBL; U86919; AAB57977.1; -.  
DR NON TER 1 1  
FT NON TER 110 110  
SQ SEQUENCE 110 AA; 11856 MW; 6A59CB25FC3DB780 CRC64;

Query Match 100.0%; Score 19; DB 12; Length 110;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 9 VAEF 12

RESULT 73  
ID Q88EG3 PRELIMINARY; PRT; 110 AA.  
AC Q88EG3;  
DT 01-JUN-2003 (TREMBLrel. 24, Created)  
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE YCII-related domain protein.  
GN PP4502.  
OS Pseudomonas putida (strain KT2440).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=160486;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22423060; PubMed=12534463;  
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,  
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,  
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,  
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,  
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,  
RA Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,  
RA Lauber J., Stjepandic D., Hohlseil J., Straetz M., Heim S.,  
RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,  
RA Fraser C.M.;  
RT "Complete genome sequence and comparative analysis of the  
RT metabolically versatile Pseudomonas putida KT2440.";  
RL Environ. Microbiol. 4:799-808(2002).  
DR EMBL; AF016790; AAN70076.1; -.  
DR TIGR; PF4502; -.  
DR InterPro; IPR005545; YCII.  
DR Pfam; PF03795; YCII; 1.  
KW Complete proteome.  
SQ SEQUENCE 110 AA; 11775 MW; 774B31C0D925A01E CRC64;

Query Match 100.0%; Score 19; DB 16; Length 110;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 73 VAEF 76

RESULT 74  
ID Q8U296 PRELIMINARY; PRT; 111 AA.  
AC Q8U296;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hypothetical protein PF0943.  
GN PF0943.  
OS Pyrococcus furiosus.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=2261;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
RT "The complete sequence of the Pyrococcus furiosus genome.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE010208; AAL81067.1; -.  
DR InterPro; IPR007842; HEPN.  
DR Pfam; PF05168; DUF712; 1.  
DR PROSITE; PS0910; HEPN; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 111 AA; 13379 MW; C6A005A6B2B6CA4E CRC64;

Query Match 100.0%; Score 19; DB 17; Length 111;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 48 VAEF 51

RESULT 75  
ID Q86PL2 PRELIMINARY; PRT; 112 AA.  
AC Q86PL2;  
DT 01-JUN-2003 (TREMBLrel. 24, Created)  
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Nuclear receptor NHR-1 (Fragment).  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Robinson-Rechavi M., Maina C.V., Gissendanner C.R., Laudet V.,  
RA Sluder A.;  
RT "Explosive lineage-specific expansion of the orphan nuclear receptor  
RT HNF4 in nematodes.";  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY204162; AA039166.1; -.  
DR GO; GO:0004872; P:receptor activity; IEA.  
DR InterPro; IPR008946; Str\_ncl\_receptor.  
KW Receptor.  
FT NON TER 1 1  
SQ SEQUENCE 112 AA; 12826 MW; 44FC60A4E3744791 CRC64;

Query Match 100.0%; Score 19; DB 5; Length 112;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 111

b 13 VAEF 16

earch completed: May 24, 2004, 17:41:04  
ob time : 29.5714 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM protein - protein search, using sw model

run on: May 24, 2004, 17:38:13 ; Search time 12 Seconds  
(without alignments)  
17.209 Million cell updates/sec

title: US-09-594-978A-2

perfect score: 19

sequence: 1 VAEF 4

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

database :

1: /cgn2\_6/prodata/2/iaa/5A COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5A COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/5A COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/5A COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/5A COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.ppep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

result	No.	Score	Query Match	Length	ID	Description
1	19	100.0	7	4	US-09-724-566A-78	Sequence 78, Appl
2	19	100.0	8	4	US-09-724-566A-81	Sequence 81, Appl
3	19	100.0	9	4	US-08-197-484-86	Sequence 86, Appl
4	19	100.0	9	4	US-08-197-484-145	Sequence 145, Appl
5	19	100.0	9	4	US-09-724-566A-73	Sequence 73, Appl
6	19	100.0	9	5	PCT-US95-02121-86	Sequence 86, Appl
7	19	100.0	9	5	PCT-US95-02121-145	Sequence 145, Appl
8	19	100.0	14	4	US-09-724-566A-72	Sequence 72, Appl
9	19	100.0	14	4	US-09-724-566A-97	Sequence 97, Appl
10	19	100.0	28	2	US-08-733-825-3	Sequence 3, Appl
11	19	100.0	35	1	US-08-487-890A-69	Sequence 69, Appl
12	19	100.0	35	1	US-08-468-763-1	Sequence 1, Appl
13	19	100.0	35	2	US-08-393-996A-1	Sequence 69, Appl
14	19	100.0	35	2	US-08-478-435-69	Sequence 69, Appl
15	19	100.0	35	2	US-08-337-483-69	Sequence 69, Appl
16	19	100.0	35	2	US-08-478-373-69	Sequence 69, Appl
17	19	100.0	35	3	US-08-474-671-69	Sequence 69, Appl
18	19	100.0	35	3	US-08-483-577A-69	Sequence 69, Appl
19	19	100.0	35	3	US-08-897-438-69	Sequence 69, Appl
20	19	100.0	35	4	US-08-637-654-69	Sequence 69, Appl
21	19	100.0	35	4	US-08-649-518-69	Sequence 69, Appl
22	19	100.0	68	4	US-09-393-634-80	Sequence 80, Appl
23	19	100.0	69	4	US-09-489-039A-13555	Sequence 13555, A
24	19	100.0	71	4	US-09-328-352-7111	Sequence 7111, Appl
25	19	100.0	73	1	US-08-321-071A-10	Sequence 10, Appl
26	19	100.0	74	4	US-08-894-139-10	Sequence 10, Appl
27	19	100.0	88	4	US-09-732-210-1412	Sequence 1412, Appl

19	100.0	93	1	US-07-612-674-12	Sequence 12, Appl
19	100.0	98	4	US-09-543-681A-5816	Sequence 5816, Appl
19	100.0	100	2	US-08-464-517-8	Sequence 8, Appl
19	100.0	100	2	US-08-246-361A-8	Sequence 8, Appl
19	100.0	100	2	US-08-463-772-8	Sequence 8, Appl
19	100.0	100	5	PCT-US93-05000-8	Sequence 8, Appl
19	100.0	101	2	US-08-580-988A-21	Sequence 21, Appl
19	100.0	103	4	US-09-540-236-3612	Sequence 3612, Appl
19	100.0	106	2	US-08-464-517-25	Sequence 25, Appl
19	100.0	106	2	US-08-246-361A-25	Sequence 25, Appl
19	100.0	106	3	US-08-463-772-25	Sequence 25, Appl
19	100.0	106	4	US-09-489-039A-8721	Sequence 8721, Appl
19	100.0	106	5	PCT-US93-05000-25	Sequence 25, Appl
19	100.0	115	4	US-09-489-039A-9915	Sequence 9915, Appl
19	100.0	122	4	US-09-252-991A-21614	Sequence 21614, A
19	100.0	123	4	US-09-489-039A-7200	Sequence 7200, Appl
19	100.0	129	4	US-08-529-055-53	Sequence 53, Appl
19	100.0	143	4	US-09-489-039A-12835	Sequence 12835, A
19	100.0	145	4	US-09-198-452A-167	Sequence 167, Appl
19	100.0	145	4	US-09-134-000C-4292	Sequence 4292, Appl
19	100.0	150	2	US-08-460-694-3	Sequence 3, Appl
19	100.0	150	3	US-08-460-744-3	Sequence 3, Appl
19	100.0	150	3	US-07-667-711B-3	Sequence 3, Appl
19	100.0	152	4	US-08-679-493A-208	Sequence 208, Appl
19	100.0	153	4	US-09-134-000C-4087	Sequence 4087, Appl
19	100.0	156	4	US-09-252-991A-20805	Sequence 20805, A
19	100.0	164	4	US-09-370-838-203	Sequence 203, Appl
19	100.0	166	4	US-09-540-236-2222	Sequence 2222, Appl
19	100.0	172	1	US-08-471-058-16	Sequence 16, Appl
19	100.0	172	1	US-08-690-095-3	Sequence 3, Appl
19	100.0	172	3	US-08-471-057-16	Sequence 16, Appl
19	100.0	172	3	US-09-113-789-3	Sequence 3, Appl
19	100.0	172	4	US-08-470-865-16	Sequence 16, Appl
19	100.0	173	1	US-08-193-977-4	Sequence 4, Appl
19	100.0	174	4	US-09-489-039A-9869	Sequence 9869, Appl
19	100.0	175	2	US-08-737-980-2	Sequence 2, Appl
19	100.0	180	4	US-09-328-352-6739	Sequence 6739, Appl
19	100.0	185	4	US-09-252-991A-24341	Sequence 24341, A
19	100.0	189	4	US-09-543-681A-7313	Sequence 7313, Appl
19	100.0	190	1	US-08-816-241-1	Sequence 1, Appl
19	100.0	190	3	US-09-128-395-1	Sequence 1, Appl
19	100.0	194	4	US-09-134-000C-3800	Sequence 3800, Appl
19	100.0	199	3	US-09-232-200-85	Sequence 85, Appl
19	100.0	199	4	US-09-232-197-85	Sequence 85, Appl
19	100.0	199	4	US-09-233-201-85	Sequence 85, Appl
19	100.0	199	4	US-09-233-195-85	Sequence 85, Appl
19	100.0	203	4	US-09-107-532A-4612	Sequence 4612, Appl
19	100.0	205	2	US-08-684-024-1	Sequence 1, Appl
19	100.0	205	2	US-08-684-024-6	Sequence 6, Appl
19	100.0	205	2	US-08-684-024-7	Sequence 7, Appl
19	100.0	205	3	US-09-145-868-1	Sequence 1, Appl
19	100.0	205	3	US-09-145-868-6	Sequence 6, Appl
19	100.0	205	3	US-09-145-868-7	Sequence 7, Appl
19	100.0	208	4	US-09-134-000C-3408	Sequence 3408, Appl
19	100.0	212	4	US-09-489-039A-10557	Sequence 10557, A
19	100.0	213	3	US-09-232-191-17	Sequence 17, Appl
19	100.0	213	3	US-09-232-200-17	Sequence 17, Appl
19	100.0	213	4	US-09-232-197-17	Sequence 17, Appl
19	100.0	213	4	US-09-233-201-17	Sequence 17, Appl
19	100.0	213	4	US-09-253-991A-30586	Sequence 30586, A
19	100.0	222	4	US-09-328-352-8105	Sequence 8105, Appl
19	100.0	225	4	US-09-489-847-155	Sequence 155, Appl
19	100.0	226	4	US-09-328-352-7363	Sequence 7363, Appl
19	100.0	228	4	US-09-455-960-8	Sequence 8, Appl
19	100.0	231	4	US-09-328-352-7924	Sequence 7924, Appl
19	100.0	231	4	US-09-134-000C-6021	Sequence 6021, Appl
19	100.0	231	4	US-09-489-039A-8897	Sequence 8897, Appl
19	100.0	234	3	US-09-248-335-34	Sequence 34, Appl
19	100.0	234	4	US-09-107-532A-8523	Sequence 8523, Appl
19	100.0	236	4	US-09-605-858-34	Sequence 34, Appl
19	100.0	238	4	US-09-605-858-35	Sequence 35, Appl
19	100.0	238	4	US-09-598-401C-68	Sequence 68, Appl

101	19	100.0	241	4	US-09-134-001C-5598	Sequence 5598, Ap	174	19	100.0	337	4	US-09-371-671B-2	Sequence 2, Appli
102	19	100.0	241	4	US-09-489-039A-7795	Sequence 7795, Ap	175	19	100.0	338	4	US-09-328-352-7151	Sequence 7151, Ap
103	19	100.0	243	4	US-09-134-001C-3587	Sequence 3587, Ap	176	19	100.0	339	3	US-09-345-468-3	Sequence 3, Appli
104	19	100.0	246	4	US-09-252-991A-30976	Sequence 30976, A	177	19	100.0	339	4	US-09-414-453A-3	Sequence 3, Appli
105	19	100.0	247	4	US-09-230-196-15	Sequence 15, Appl	178	19	100.0	339	4	US-09-134-000C-5209	Sequence 5209, Ap
106	19	100.0	247	4	US-09-372-422A-48	Sequence 48, Appl	179	19	100.0	342	4	US-09-543-681A-4207	Sequence 4207, Ap
107	19	100.0	249	3	US-09-154-083-2	Sequence 2, Appli	180	19	100.0	345	4	US-09-107-532A-3849	Sequence 3849, Ap
108	19	100.0	249	3	US-09-345-468-9	Sequence 9, Appli	181	19	100.0	346	4	US-09-266-965-107	Sequence 107, App
109	19	100.0	249	4	US-09-372-422A-22	Sequence 22, Appl	182	19	100.0	347	4	US-09-134-001C-3298	Sequence 3298, Ap
110	19	100.0	249	4	US-09-414-453A-9	Sequence 9, Appli	183	19	100.0	350	4	US-09-655-270A-17	Sequence 17, Appl
111	19	100.0	249	4	US-09-252-991A-22610	Sequence 22610, A	184	19	100.0	350	4	US-09-651-941-21	Sequence 21, Appl
112	19	100.0	250	1	US-08-234-939-2	Sequence 2, Appli	185	19	100.0	350	4	US-09-955-597-21	Sequence 21, Appl
113	19	100.0	250	1	US-08-558-865-2	Sequence 2, Appli	186	19	100.0	353	4	US-09-252-991A-27528	Sequence 27528, A
114	19	100.0	250	3	US-08-654-025-2	Sequence 2, Appli	187	19	100.0	356	4	US-09-328-352-7671	Sequence 7671, Ap
115	19	100.0	250	3	US-08-654-025-7	Sequence 7, Appli	188	19	100.0	357	4	US-09-540-236-3808	Sequence 3808, Ap
116	19	100.0	251	4	US-09-252-991A-23816	Sequence 23816, A	189	19	100.0	363	4	US-09-252-991A-30015	Sequence 30015, A
117	19	100.0	253	4	US-09-252-991A-17352	Sequence 17352, A	190	19	100.0	364	4	US-09-543-681A-7525	Sequence 7525, Ap
118	19	100.0	253	4	US-09-107-532A-5072	Sequence 5072, Ap	191	19	100.0	368	4	US-09-252-991A-24621	Sequence 24621, A
119	19	100.0	254	4	US-09-372-422A-34	Sequence 34, Appl	192	19	100.0	368	4	US-09-489-039A-9556	Sequence 9556, Ap
120	19	100.0	257	2	US-08-506-340A-3	Sequence 3, Appli	193	19	100.0	370	4	US-09-540-224-4	Sequence 4, Appli
121	19	100.0	257	4	US-09-543-681A-6811	Sequence 6811, Ap	194	19	100.0	370	4	US-09-564-595D-53	Sequence 53, Appli
122	19	100.0	259	4	US-09-543-681A-6390	Sequence 6390, Ap	195	19	100.0	370	4	US-09-808-972-4	Sequence 4, Appli
123	19	100.0	259	4	US-09-489-039A-11936	Sequence 11936, A	196	19	100.0	371	2	US-08-837-593-6	Sequence 6, Appli
124	19	100.0	261	4	US-09-252-991A-23795	Sequence 23795, A	197	19	100.0	372	4	US-09-800-729-213	Sequence 213, App
125	19	100.0	262	4	US-09-252-991A-19738	Sequence 19738, A	198	19	100.0	374	4	US-09-489-039A-12678	Sequence 12678, A
126	19	100.0	263	4	US-09-800-729-88	Sequence 88, Appl	199	19	100.0	388	4	US-09-489-039A-8170	Sequence 8170, Ap
127	19	100.0	269	1	US-08-447-554-5	Sequence 5, Appli	200	19	100.0	390	4	US-09-933-313B-8	Sequence 8, Appli
128	19	100.0	269	1	US-08-468-763-17	Sequence 17, Appl	201	19	100.0	390	4	US-08-311-731A-332	Sequence 332, App
129	19	100.0	269	1	US-08-448-160-5	Sequence 5, Appli	202	19	100.0	390	4	US-09-543-681A-6415	Sequence 6415, Ap
130	19	100.0	269	2	US-08-393-966A-17	Sequence 17, Appl	203	19	100.0	390	4	US-09-489-039A-13547	Sequence 13547, A
131	19	100.0	270	4	US-09-252-991A-29192	Sequence 29192, A	204	19	100.0	391	2	US-08-928-692-26	Sequence 26, Appl
132	19	100.0	271	4	US-09-107-532A-5071	Sequence 5071, Ap	205	19	100.0	391	4	US-09-339-972-26	Sequence 26, Appl
133	19	100.0	273	4	US-09-489-039A-9190	Sequence 9190, Ap	206	19	100.0	394	4	US-09-934-903-4	Sequence 4, Appli
134	19	100.0	277	1	US-08-400-413-1	Sequence 1, Appli	207	19	100.0	397	4	US-09-252-991A-22235	Sequence 22235, A
135	19	100.0	280	1	US-08-595-559-3	Sequence 3, Appli	208	19	100.0	398	4	US-09-242-859A-4	Sequence 4, Appli
136	19	100.0	281	4	US-09-314-701-58	Sequence 58, Appl	209	19	100.0	398	4	US-08-242-859A-8	Sequence 8, Appli
137	19	100.0	282	4	US-09-134-000C-3572	Sequence 3572, Ap	210	19	100.0	398	4	US-09-252-991A-17379	Sequence 17379, A
138	19	100.0	283	4	US-08-956-171E-5203	Sequence 5203, Ap	211	19	100.0	400	2	US-08-733-825-2	Sequence 2, Appli
139	19	100.0	288	4	US-09-252-991A-18721	Sequence 18721, A	212	19	100.0	400	3	US-09-264-097-6	Sequence 6, Appli
140	19	100.0	290	4	US-09-743-847-2	Sequence 2, Appli	213	19	100.0	403	4	US-09-540-236-2573	Sequence 2573, Ap
141	19	100.0	291	4	US-09-252-991A-25517	Sequence 25517, A	214	19	100.0	405	3	US-09-232-200-63	Sequence 63, Appl
142	19	100.0	291	4	US-09-489-039A-9578	Sequence 9578, Ap	215	19	100.0	405	4	US-09-232-197-63	Sequence 63, Appl
143	19	100.0	294	4	US-09-523-263B-20	Sequence 20, Appl	216	19	100.0	405	4	US-09-232-201-63	Sequence 20, Appl
144	19	100.0	295	4	US-09-543-681A-5660	Sequence 5660, Ap	217	19	100.0	405	4	US-09-232-201-63	Sequence 20, Appl
145	19	100.0	297	4	US-09-489-039A-9087	Sequence 9087, Ap	218	19	100.0	405	4	US-09-540-715A-20	Sequence 20, Appl
146	19	100.0	299	4	US-09-314-701-46	Sequence 46, Appl	219	19	100.0	405	4	US-09-232-195-63	Sequence 63, Appl
147	19	100.0	299	4	US-09-393-634-35	Sequence 35, Appl	220	19	100.0	406	4	US-09-543-681A-7962	Sequence 7962, Ap
148	19	100.0	301	4	US-09-489-039A-11964	Sequence 11964, A	221	19	100.0	410	2	US-08-723-415B-11	Sequence 10, Appl
149	19	100.0	311	4	US-09-614-912-198	Sequence 198, App	222	19	100.0	410	2	US-08-723-415B-11	Sequence 11, Appl
150	19	100.0	317	4	US-09-489-077A-17	Sequence 17, Appl	223	19	100.0	410	2	US-08-428-331-2	Sequence 2, Appli
151	19	100.0	318	2	US-08-872-719-2	Sequence 2, Appli	224	19	100.0	410	2	US-08-602-846-2	Sequence 2, Appli
152	19	100.0	318	3	US-08-957-302A-12	Sequence 12, Appl	225	19	100.0	410	3	US-09-078-596-2	Sequence 2, Appli
153	19	100.0	318	3	US-09-336-890-2	Sequence 2, Appli	226	19	100.0	410	3	US-09-189-627A-11	Sequence 10, Appl
154	19	100.0	318	3	US-09-542-403-12	Sequence 12, Appl	227	19	100.0	410	3	US-09-189-627A-11	Sequence 10, Appl
155	19	100.0	318	4	US-09-668-499-2	Sequence 2, Appli	228	19	100.0	410	4	US-09-710-861-11	Sequence 10, Appl
156	19	100.0	318	4	US-09-976-594-938	Sequence 938, App	229	19	100.0	410	4	US-09-252-991A-24839	Sequence 24839, A
157	19	100.0	319	2	US-08-793-927-4	Sequence 4, Appli	230	19	100.0	410	4	US-09-540-236-3549	Sequence 3549, Ap
158	19	100.0	319	3	US-09-345-468-5	Sequence 5, Appli	231	19	100.0	411	4	US-09-252-991A-17176	Sequence 17176, A
159	19	100.0	319	4	US-09-414-453A-5	Sequence 5, Appli	232	19	100.0	414	4	US-09-252-991A-5101	Sequence 5101, Ap
160	19	100.0	321	4	US-09-440-597-4	Sequence 4, Appli	233	19	100.0	415	4	US-09-134-001C-5101	Sequence 5101, Ap
161	19	100.0	321	4	US-09-252-991A-18807	Sequence 18807, A	234	19	100.0	415	4	US-09-252-991A-31684	Sequence 31684, A
162	19	100.0	323	4	US-09-252-991A-17111	Sequence 17111, A	235	19	100.0	416	4	US-09-252-991A-26713	Sequence 26713, A
163	19	100.0	323	4	US-09-543-681A-6957	Sequence 6957, Ap	236	19	100.0	424	4	US-09-328-352-4187	Sequence 4187, Ap
164	19	100.0	323	4	US-09-489-039A-7408	Sequence 7408, Ap	237	19	100.0	424	4	US-09-489-039A-9628	Sequence 9628, Ap
165	19	100.0	324	4	US-09-328-352-4636	Sequence 4636, Ap	238	19	100.0	425	4	US-09-489-039A-8386	Sequence 8386, Ap
166	19	100.0	326	4	US-09-230-196-20	Sequence 20, Appl	239	19	100.0	431	4	US-09-134-001C-4257	Sequence 4257, Ap
167	19	100.0	327	4	US-09-489-039A-9512	Sequence 9512, Ap	240	19	100.0	431	4	US-09-252-991A-24878	Sequence 24878, A
168	19	100.0	327	4	US-09-252-991A-30065	Sequence 30065, A	241	19	100.0	432	1	US-08-522-166-8	Sequence 8, Appli
169	19	100.0	328	4	US-09-489-039A-9306	Sequence 9306, Ap	242	19	100.0	432	1	US-08-488-382A-8	Sequence 8, Appli
170	19	100.0	331	4	US-09-489-039A-9639	Sequence 9639, Ap	243	19	100.0	432	2	US-08-480-912-8	Sequence 8, Appli
171	19	100.0	335	1	US-08-202-054-2	Sequence 2, Appli	244	19	100.0	435	4	US-09-252-991A-30562	Sequence 30562, A
172	19	100.0	335	1	US-08-446-923-2	Sequence 2, Appli	245	19	100.0	436	4	US-09-161-994A-9	Sequence 9, Appli
173	19	100.0	335	4	US-09-482-273-118	Sequence 118, App	246	19	100.0	436	4	US-09-134-000C-3950	Sequence 3950, Ap

247	19	100.0	436	6	5405943-4	Patent No. 5405943	320	19	100.0	485	4	US-09-540-715A-18	Sequence 18, Appl
248	19	100.0	437	4	US-09-252-991A-25332	Sequence 25332, A	321	19	100.0	485	4	US-09-540-715A-19	Sequence 19, Appl
249	19	100.0	440	4	US-09-252-991A-22487	Sequence 22487, A	322	19	100.0	485	4	US-09-769-864-1	Sequence 1, Appl
250	19	100.0	444	4	US-09-252-991A-18256	Sequence 18256, A	323	19	100.0	485	4	US-09-769-864-2	Sequence 2, Appl
251	19	100.0	444	4	US-09-252-991A-18565	Sequence 18565, A	324	19	100.0	485	4	US-09-769-864-6	Sequence 6, Appl
252	19	100.0	445	4	US-09-252-991A-28986	Sequence 28986, A	325	19	100.0	485	4	US-09-769-864-7	Sequence 7, Appl
253	19	100.0	445	4	US-09-107-532A-7232	Sequence 7232, Ap	326	19	100.0	485	4	US-09-769-864-8	Sequence 8, Appl
254	19	100.0	446	4	US-09-328-352-6076	Sequence 6076, Ap	327	19	100.0	486	4	US-09-381-687-4	Sequence 4, Appl
255	19	100.0	449	4	US-09-134-000C-5444	Sequence 5444, Ap	328	19	100.0	493	3	US-08-996-441B-70	Sequence 70, Appl
256	19	100.0	450	4	US-09-543-681A-7260	Sequence 7260, Ap	329	19	100.0	493	3	US-08-993-722A-70	Sequence 70, Appl
257	19	100.0	451	4	US-09-134-000C-6421	Sequence 6421, Ap	330	19	100.0	493	3	US-08-993-170A-70	Sequence 70, Appl
258	19	100.0	455	4	US-09-252-991A-19936	Sequence 19936, A	331	19	100.0	493	3	US-08-993-775B-70	Sequence 70, Appl
259	19	100.0	456	4	US-09-252-991A-31884	Sequence 31884, A	332	19	100.0	493	3	US-09-177-349-5	Sequence 5, Appl
260	19	100.0	457	4	US-09-252-991A-28824	Sequence 28824, A	333	19	100.0	493	4	US-09-328-352-6753	Sequence 6753, Ap
261	19	100.0	467	4	US-09-252-991A-28824	Sequence 28824, A	334	19	100.0	493	4	US-09-427-770-70	Sequence 70, Appl
262	19	100.0	468	4	US-09-252-991A-18608	Sequence 18608, A	335	19	100.0	493	4	US-09-427-769-70	Sequence 70, Appl
263	19	100.0	469	4	US-09-543-681A-5423	Sequence 5423, Ap	336	19	100.0	493	4	US-09-540-236-2869	Sequence 2869, Ap
264	19	100.0	474	4	US-09-252-991A-27524	Sequence 27524, A	337	19	100.0	495	3	US-09-179-966-1	Sequence 1, Appl
265	19	100.0	475	4	US-09-252-991A-32806	Sequence 32806, A	338	19	100.0	495	3	US-08-980-994-1	Sequence 1, Appl
266	19	100.0	479	4	US-09-328-352-5396	Sequence 5396, Ap	339	19	100.0	504	4	US-09-252-991A-20317	Sequence 20317, A
267	19	100.0	485	2	US-08-446-803-1	Sequence 1, Appl	340	19	100.0	506	3	US-09-232-181-9	Sequence 9, Appl
268	19	100.0	485	2	US-08-446-803-2	Sequence 2, Appl	341	19	100.0	506	3	US-09-232-200-9	Sequence 9, Appl
269	19	100.0	485	2	US-08-861-837-1	Sequence 1, Appl	342	19	100.0	506	3	US-09-232-200-95	Sequence 95, Appl
270	19	100.0	485	2	US-08-861-837-2	Sequence 2, Appl	343	19	100.0	506	4	US-09-232-197-9	Sequence 9, Appl
271	19	100.0	485	2	US-08-600-908A-12	Sequence 12, Appl	344	19	100.0	506	4	US-09-232-197-95	Sequence 95, Appl
272	19	100.0	485	3	US-08-683-838A-12	Sequence 12, Appl	345	19	100.0	506	4	US-09-232-201-9	Sequence 9, Appl
273	19	100.0	485	3	US-08-600-656-1	Sequence 1, Appl	346	19	100.0	506	4	US-09-232-201-95	Sequence 95, Appl
274	19	100.0	485	3	US-08-600-656-2	Sequence 2, Appl	347	19	100.0	506	4	US-09-232-185-9	Sequence 9, Appl
275	19	100.0	485	3	US-08-600-656-7	Sequence 7, Appl	348	19	100.0	506	4	US-09-232-195-95	Sequence 95, Appl
276	19	100.0	485	3	US-09-170-670-1	Sequence 1, Appl	349	19	100.0	516	4	US-09-986-676A-2	Sequence 2, Appl
277	19	100.0	485	3	US-09-170-670-2	Sequence 2, Appl	350	19	100.0	522	4	US-09-252-991A-30451	Sequence 30451, A
278	19	100.0	485	3	US-09-170-670-6	Sequence 6, Appl	351	19	100.0	524	2	US-08-928-692-12	Sequence 12, Appl
279	19	100.0	485	3	US-09-170-670-7	Sequence 7, Appl	352	19	100.0	524	3	US-08-957-302A-2	Sequence 2, Appl
280	19	100.0	485	3	US-09-170-670-8	Sequence 8, Appl	353	19	100.0	524	3	US-09-543-403-2	Sequence 2, Appl
281	19	100.0	485	3	US-09-193-068-1	Sequence 1, Appl	354	19	100.0	524	4	US-09-333-972-12	Sequence 12, Appl
282	19	100.0	485	3	US-09-193-068-2	Sequence 2, Appl	355	19	100.0	532	4	US-09-252-991A-27288	Sequence 27288, A
283	19	100.0	485	3	US-09-193-068-6	Sequence 6, Appl	356	19	100.0	536	4	US-09-107-532A-6930	Sequence 6930, Ap
284	19	100.0	485	3	US-09-193-068-7	Sequence 7, Appl	357	19	100.0	538	4	US-09-489-039A-13518	Sequence 13518, A
285	19	100.0	485	3	US-09-193-068-8	Sequence 8, Appl	358	19	100.0	540	4	US-09-621-451-2	Sequence 2, Appl
286	19	100.0	485	3	US-09-183-412-1	Sequence 1, Appl	359	19	100.0	540	4	US-10-223-355-2	Sequence 2, Appl
287	19	100.0	485	3	US-09-183-412-2	Sequence 2, Appl	360	19	100.0	544	4	US-09-328-352-4446	Sequence 4446, Ap
288	19	100.0	485	3	US-09-183-412-6	Sequence 6, Appl	361	19	100.0	547	3	US-08-483-577A-149	Sequence 149, App
289	19	100.0	485	3	US-09-183-412-7	Sequence 7, Appl	362	19	100.0	547	3	US-08-897-438-149	Sequence 149, App
290	19	100.0	485	3	US-09-183-412-8	Sequence 8, Appl	363	19	100.0	547	4	US-08-649-518-149	Sequence 149, App
291	19	100.0	485	3	US-09-264-097-5	Sequence 5, Appl	364	19	100.0	550	4	US-09-396-478A-2	Sequence 2, Appl
292	19	100.0	485	3	US-09-264-097-7	Sequence 7, Appl	365	19	100.0	552	4	US-09-489-039A-13735	Sequence 13735, A
293	19	100.0	485	3	US-09-354-191A-1	Sequence 1, Appl	366	19	100.0	557	4	US-09-134-001C-5569	Sequence 5569, Ap
294	19	100.0	485	3	US-09-354-191A-2	Sequence 2, Appl	367	19	100.0	557	4	US-09-489-039A-13795	Sequence 13795, Ap
295	19	100.0	485	3	US-09-354-191A-7	Sequence 7, Appl	368	19	100.0	558	4	US-09-489-039A-13795	Sequence 13795, Ap
296	19	100.0	485	4	US-09-291-022A-13	Sequence 13, Appl	369	19	100.0	566	4	US-09-252-991A-27678	Sequence 27678, A
297	19	100.0	485	4	US-09-291-022A-18	Sequence 18, Appl	370	19	100.0	566	4	US-09-513-783A-142	Sequence 142, App
298	19	100.0	485	4	US-09-291-022A-19	Sequence 19, Appl	371	19	100.0	567	2	US-08-504-459-4	Sequence 4, Appl
299	19	100.0	485	4	US-09-290-734-1	Sequence 1, Appl	372	19	100.0	567	2	US-08-504-459-6	Sequence 6, Appl
300	19	100.0	485	4	US-09-290-734-2	Sequence 2, Appl	373	19	100.0	571	3	US-08-503-326A-6	Sequence 6, Appl
301	19	100.0	485	4	US-09-290-734-6	Sequence 6, Appl	374	19	100.0	571	4	US-09-689-914A-6	Sequence 6, Appl
302	19	100.0	485	4	US-09-290-734-7	Sequence 7, Appl	375	19	100.0	571	4	US-09-689-914A-6	Sequence 6, Appl
303	19	100.0	485	4	US-09-290-734-8	Sequence 8, Appl	376	19	100.0	571	4	US-09-689-914A-6	Sequence 6, Appl
304	19	100.0	485	4	US-09-290-734-24	Sequence 24, Appl	377	19	100.0	573	4	US-09-252-991A-18744	Sequence 18744, A
305	19	100.0	485	4	US-09-290-734-26	Sequence 26, Appl	378	19	100.0	573	4	US-09-328-352-4675	Sequence 4675, Ap
306	19	100.0	485	4	US-09-636-252A-12	Sequence 12, Appl	379	19	100.0	580	4	US-09-198-452A-332	Sequence 332, App
307	19	100.0	485	4	US-09-417-355A-5	Sequence 5, Appl	380	19	100.0	587	4	US-08-635-552A-3	Sequence 3, Appl
308	19	100.0	485	4	US-09-381-687-1	Sequence 1, Appl	381	19	100.0	587	4	US-08-675-499A-4	Sequence 4, Appl
309	19	100.0	485	4	US-09-381-687-2	Sequence 2, Appl	382	19	100.0	587	4	US-08-812-008-4	Sequence 4, Appl
310	19	100.0	485	4	US-09-381-687-3	Sequence 3, Appl	383	19	100.0	589	4	US-09-643-657-14	Sequence 14, Appl
311	19	100.0	485	4	US-09-381-687-5	Sequence 5, Appl	384	19	100.0	591	2	US-08-736-770-5	Sequence 5, Appl
312	19	100.0	485	4	US-09-545-586-1	Sequence 1, Appl	385	19	100.0	591	4	US-09-643-657-5	Sequence 5, Appl
313	19	100.0	485	4	US-09-545-586-2	Sequence 2, Appl	386	19	100.0	591	4	US-09-643-657-15	Sequence 15, Appl
314	19	100.0	485	4	US-09-545-586-6	Sequence 6, Appl	387	19	100.0	597	4	US-09-252-991A-23152	Sequence 23152, Ap
315	19	100.0	485	4	US-09-545-586-7	Sequence 7, Appl	388	19	100.0	597	4	US-09-540-236-2805	Sequence 2805, Ap
316	19	100.0	485	4	US-09-545-586-8	Sequence 8, Appl	389	19	100.0	607	2	US-08-472-534-5	Sequence 5, Appl
317	19	100.0	485	4	US-09-545-586-24	Sequence 24, Appl	390	19	100.0	608	2	US-08-736-770-1	Sequence 1, Appl
318	19	100.0	485	4	US-09-545-586-26	Sequence 26, Appl	391	19	100.0	620	4	US-08-637-670-40	Sequence 40, Appl
319	19	100.0	485	4	US-09-540-715A-13	Sequence 13, Appl	392	19	100.0	623	3	US-09-041-991A-6	Sequence 6, Appl

393	19	100.0	623	4	US-09-608-533A-6	Sequence 6, Appli	466	19	100.0	651	3	US-08-993-170A-56	Sequence 56, Appl
394	19	100.0	625	4	US-09-661-322A-48	Sequence 48, Appl	467	19	100.0	651	3	US-08-993-170A-58	Sequence 58, Appl
395	19	100.0	626	4	US-09-252-991A-20268	Sequence 20268, A	468	19	100.0	651	3	US-08-993-775B-52	Sequence 52, Appl
396	19	100.0	631	1	US-08-487-890A-111	Sequence 111, App	469	19	100.0	651	3	US-08-993-775B-56	Sequence 56, Appl
397	19	100.0	631	2	US-08-478-435-111	Sequence 111, App	470	19	100.0	651	3	US-08-993-775B-58	Sequence 58, Appl
398	19	100.0	631	2	US-08-337-483-111	Sequence 111, App	471	19	100.0	651	4	US-09-427-770-52	Sequence 52, Appl
399	19	100.0	631	2	US-08-478-373-111	Sequence 111, App	472	19	100.0	651	4	US-09-427-770-56	Sequence 56, Appl
400	19	100.0	631	3	US-08-474-671-111	Sequence 111, App	473	19	100.0	651	4	US-09-427-770-58	Sequence 58, Appl
401	19	100.0	631	3	US-08-483-577A-111	Sequence 111, App	474	19	100.0	651	4	US-09-427-769-52	Sequence 52, Appl
402	19	100.0	631	3	US-08-897-438-111	Sequence 111, App	475	19	100.0	651	4	US-09-427-769-56	Sequence 56, Appl
403	19	100.0	631	4	US-08-637-654-111	Sequence 111, App	476	19	100.0	651	4	US-09-427-769-58	Sequence 58, Appl
404	19	100.0	631	4	US-08-649-518-111	Sequence 111, App	477	19	100.0	652	3	US-08-996-441B-2	Sequence 2, Appli
405	19	100.0	632	3	US-09-232-200-34	Sequence 34, Appl	478	19	100.0	652	3	US-08-996-441B-4	Sequence 4, Appli
406	19	100.0	632	3	US-09-232-200-35	Sequence 35, Appl	479	19	100.0	652	3	US-08-996-441B-6	Sequence 6, Appli
407	19	100.0	632	3	US-09-232-200-39	Sequence 39, Appl	480	19	100.0	652	3	US-08-996-441B-8	Sequence 8, Appli
408	19	100.0	632	4	US-09-232-197-34	Sequence 34, Appl	481	19	100.0	652	3	US-08-996-441B-10	Sequence 10, Appl
409	19	100.0	632	4	US-09-232-197-35	Sequence 35, Appl	482	19	100.0	652	3	US-08-996-441B-12	Sequence 12, Appl
410	19	100.0	632	4	US-09-232-197-39	Sequence 39, Appl	483	19	100.0	652	3	US-08-996-441B-14	Sequence 14, Appl
411	19	100.0	632	4	US-09-232-201-34	Sequence 34, Appl	484	19	100.0	652	3	US-08-996-441B-16	Sequence 16, Appl
412	19	100.0	632	4	US-09-232-201-35	Sequence 35, Appl	485	19	100.0	652	3	US-08-996-441B-18	Sequence 18, Appl
413	19	100.0	632	4	US-09-232-201-39	Sequence 39, Appl	486	19	100.0	652	3	US-08-996-441B-20	Sequence 20, Appl
414	19	100.0	632	4	US-09-232-195-34	Sequence 34, Appl	487	19	100.0	652	3	US-08-996-441B-22	Sequence 22, Appl
415	19	100.0	632	4	US-09-232-195-35	Sequence 35, Appl	488	19	100.0	652	3	US-08-996-441B-24	Sequence 24, Appl
416	19	100.0	632	4	US-09-232-195-39	Sequence 39, Appl	489	19	100.0	652	3	US-08-996-441B-26	Sequence 26, Appl
417	19	100.0	635	4	US-09-107-532A-7135	Sequence 7135, Ap	490	19	100.0	652	3	US-08-996-441B-28	Sequence 28, Appl
418	19	100.0	638	4	US-09-489-039A-8996	Sequence 8996, Ap	491	19	100.0	652	3	US-08-996-441B-30	Sequence 30, Appl
419	19	100.0	639	4	US-09-328-352-4979	Sequence 4979, Ap	492	19	100.0	652	3	US-08-996-441B-32	Sequence 32, Appl
420	19	100.0	640	4	US-09-253-991A-23007	Sequence 23007, A	493	19	100.0	652	3	US-08-996-441B-34	Sequence 34, Appl
421	19	100.0	641	4	US-09-543-681A-6258	Sequence 6258, Ap	494	19	100.0	652	3	US-08-996-441B-36	Sequence 36, Appl
422	19	100.0	643	3	US-09-232-200-27	Sequence 27, Appl	495	19	100.0	652	3	US-08-996-441B-38	Sequence 38, Appl
423	19	100.0	643	3	US-09-232-200-41	Sequence 41, Appl	496	19	100.0	652	3	US-08-996-441B-40	Sequence 40, Appl
424	19	100.0	643	3	US-09-232-200-42	Sequence 42, Appl	497	19	100.0	652	3	US-08-996-441B-42	Sequence 42, Appl
425	19	100.0	643	3	US-09-232-200-45	Sequence 45, Appl	498	19	100.0	652	3	US-08-996-441B-44	Sequence 44, Appl
426	19	100.0	643	3	US-09-232-200-53	Sequence 53, Appl	499	19	100.0	652	3	US-08-996-441B-46	Sequence 46, Appl
427	19	100.0	643	3	US-09-232-200-71	Sequence 71, Appl	500	19	100.0	652	3	US-08-996-441B-48	Sequence 48, Appl
428	19	100.0	643	4	US-09-232-197-27	Sequence 27, Appl	501	19	100.0	652	3	US-08-996-441B-50	Sequence 50, Appl
429	19	100.0	643	4	US-09-232-197-41	Sequence 41, Appl	502	19	100.0	652	3	US-08-996-441B-54	Sequence 54, Appl
430	19	100.0	643	4	US-09-232-197-42	Sequence 42, Appl	503	19	100.0	652	3	US-08-996-441B-60	Sequence 60, Appl
431	19	100.0	643	4	US-09-232-197-53	Sequence 53, Appl	504	19	100.0	652	3	US-08-996-441B-62	Sequence 62, Appl
432	19	100.0	643	4	US-09-232-197-53	Sequence 53, Appl	505	19	100.0	652	3	US-08-996-441B-64	Sequence 64, Appl
433	19	100.0	643	4	US-09-232-197-71	Sequence 71, Appl	506	19	100.0	652	3	US-08-996-441B-66	Sequence 66, Appl
434	19	100.0	643	4	US-09-232-201-27	Sequence 27, Appl	507	19	100.0	652	3	US-08-996-441B-68	Sequence 68, Appl
435	19	100.0	643	4	US-09-232-201-41	Sequence 41, Appl	508	19	100.0	652	3	US-08-996-441B-98	Sequence 98, Appl
436	19	100.0	643	4	US-09-232-201-42	Sequence 42, Appl	509	19	100.0	652	3	US-08-996-441B-108	Sequence 108, App
437	19	100.0	643	4	US-09-232-201-45	Sequence 45, Appl	510	19	100.0	652	3	US-08-996-441B-110	Sequence 110, App
438	19	100.0	643	4	US-09-232-201-53	Sequence 53, Appl	511	19	100.0	652	3	US-08-996-441B-111	Sequence 111, App
439	19	100.0	643	4	US-09-232-201-71	Sequence 71, Appl	512	19	100.0	652	3	US-08-993-722A-2	Sequence 2, Appli
440	19	100.0	643	4	US-09-543-681A-6345	Sequence 6345, Ap	513	19	100.0	652	3	US-08-993-722A-4	Sequence 4, Appli
441	19	100.0	643	4	US-09-232-195-27	Sequence 27, Appl	514	19	100.0	652	3	US-08-993-722A-6	Sequence 6, Appli
442	19	100.0	643	4	US-09-232-195-41	Sequence 41, Appl	515	19	100.0	652	3	US-08-993-722A-8	Sequence 8, Appli
443	19	100.0	643	4	US-09-232-195-42	Sequence 42, Appl	516	19	100.0	652	3	US-08-993-722A-10	Sequence 10, Appl
444	19	100.0	643	4	US-09-232-195-45	Sequence 45, Appl	517	19	100.0	652	3	US-08-993-722A-12	Sequence 12, Appl
445	19	100.0	643	4	US-09-232-195-53	Sequence 53, Appl	518	19	100.0	652	3	US-08-993-722A-14	Sequence 14, Appl
446	19	100.0	643	4	US-09-232-195-71	Sequence 71, Appl	519	19	100.0	652	3	US-08-993-722A-16	Sequence 16, Appl
447	19	100.0	644	1	US-08-487-890A-6	Sequence 6, Appli	520	19	100.0	652	3	US-08-993-722A-18	Sequence 18, Appl
448	19	100.0	644	2	US-08-478-435-6	Sequence 6, Appli	521	19	100.0	652	3	US-08-993-722A-20	Sequence 20, Appl
449	19	100.0	644	2	US-08-337-483-6	Sequence 6, Appli	522	19	100.0	652	3	US-08-993-722A-22	Sequence 22, Appl
450	19	100.0	644	2	US-08-478-373-6	Sequence 6, Appli	523	19	100.0	652	3	US-08-993-722A-24	Sequence 24, Appl
451	19	100.0	644	3	US-08-474-671-6	Sequence 6, Appli	524	19	100.0	652	3	US-08-993-722A-26	Sequence 26, Appl
452	19	100.0	644	3	US-08-483-577A-6	Sequence 6, Appli	525	19	100.0	652	3	US-08-993-722A-28	Sequence 28, Appl
453	19	100.0	644	3	US-08-897-438-6	Sequence 6, Appli	526	19	100.0	652	3	US-08-993-722A-30	Sequence 30, Appl
454	19	100.0	644	4	US-08-637-654-6	Sequence 6, Appli	527	19	100.0	652	3	US-08-993-722A-32	Sequence 32, Appl
455	19	100.0	644	4	US-08-649-518-6	Sequence 6, Appli	528	19	100.0	652	3	US-08-993-722A-34	Sequence 34, Appl
456	19	100.0	647	3	US-08-483-577A-148	Sequence 148, App	529	19	100.0	652	3	US-08-993-722A-36	Sequence 36, Appl
457	19	100.0	647	3	US-08-897-438-148	Sequence 148, App	530	19	100.0	652	3	US-08-993-722A-38	Sequence 38, Appl
458	19	100.0	647	4	US-08-643-518-148	Sequence 52, App	531	19	100.0	652	3	US-08-993-722A-40	Sequence 40, Appl
459	19	100.0	651	3	US-08-996-441B-52	Sequence 52, Appl	532	19	100.0	652	3	US-08-993-722A-42	Sequence 42, Appl
460	19	100.0	651	3	US-08-996-441B-56	Sequence 56, Appl	533	19	100.0	652	3	US-08-993-722A-44	Sequence 44, Appl
461	19	100.0	651	3	US-08-996-441B-58	Sequence 58, Appl	534	19	100.0	652	3	US-08-993-722A-46	Sequence 46, Appl
462	19	100.0	651	3	US-08-993-722A-52	Sequence 52, Appl	535	19	100.0	652	3	US-08-993-722A-48	Sequence 48, Appl
463	19	100.0	651	3	US-08-993-722A-56	Sequence 56, Appl	536	19	100.0	652	3	US-08-993-722A-50	Sequence 50, Appl
464	19	100.0	651	3	US-08-993-722A-58	Sequence 58, Appl	537	19	100.0	652	3	US-08-993-722A-54	Sequence 54, Appl
465	19	100.0	651	3	US-08-993-170A-52	Sequence 52, Appl	538	19	100.0	652	3	US-08-993-722A-60	Sequence 60, Appl

539	19	100.0	652	3	US-08-993-722A-62	Sequence 62, Appl	612	19	100.0	652	3	US-08-993-775B-68	Sequence 68, Appl
540	19	100.0	652	3	US-08-993-722A-64	Sequence 64, Appl	613	19	100.0	652	3	US-08-993-775B-98	Sequence 98, Appl
541	19	100.0	652	3	US-08-993-722A-66	Sequence 66, Appl	614	19	100.0	652	3	US-08-993-775B-108	Sequence 108, Appl
542	19	100.0	652	3	US-08-993-722A-68	Sequence 68, Appl	615	19	100.0	652	3	US-08-993-775B-110	Sequence 110, Appl
543	19	100.0	652	3	US-08-993-722A-98	Sequence 98, Appl	616	19	100.0	652	3	US-08-993-775B-111	Sequence 111, Appl
544	19	100.0	652	3	US-08-993-722A-108	Sequence 108, Appl	617	19	100.0	652	4	US-09-377-466B-2	Sequence 2, Appl
545	19	100.0	652	3	US-08-993-722A-110	Sequence 110, Appl	618	19	100.0	652	4	US-09-377-466B-4	Sequence 4, Appl
546	19	100.0	652	3	US-08-993-722A-111	Sequence 111, Appl	619	19	100.0	652	4	US-09-377-466B-6	Sequence 6, Appl
547	19	100.0	652	3	US-08-993-170A-2	Sequence 2, Appl	620	19	100.0	652	4	US-09-427-770-2	Sequence 2, Appl
548	19	100.0	652	3	US-08-993-170A-4	Sequence 4, Appl	621	19	100.0	652	4	US-09-427-770-4	Sequence 4, Appl
549	19	100.0	652	3	US-08-993-170A-6	Sequence 6, Appl	622	19	100.0	652	4	US-09-427-770-6	Sequence 6, Appl
550	19	100.0	652	3	US-08-993-170A-8	Sequence 8, Appl	623	19	100.0	652	4	US-09-427-770-8	Sequence 8, Appl
551	19	100.0	652	3	US-08-993-170A-10	Sequence 10, Appl	624	19	100.0	652	4	US-09-427-770-10	Sequence 10, Appl
552	19	100.0	652	3	US-08-993-170A-12	Sequence 12, Appl	625	19	100.0	652	4	US-09-427-770-12	Sequence 12, Appl
553	19	100.0	652	3	US-08-993-170A-14	Sequence 14, Appl	626	19	100.0	652	4	US-09-427-770-14	Sequence 14, Appl
554	19	100.0	652	3	US-08-993-170A-16	Sequence 16, Appl	627	19	100.0	652	4	US-09-427-770-16	Sequence 16, Appl
555	19	100.0	652	3	US-08-993-170A-18	Sequence 18, Appl	628	19	100.0	652	4	US-09-427-770-18	Sequence 18, Appl
556	19	100.0	652	3	US-08-993-170A-20	Sequence 20, Appl	629	19	100.0	652	4	US-09-427-770-20	Sequence 20, Appl
557	19	100.0	652	3	US-08-993-170A-22	Sequence 22, Appl	630	19	100.0	652	4	US-09-427-770-22	Sequence 22, Appl
558	19	100.0	652	3	US-08-993-170A-24	Sequence 24, Appl	631	19	100.0	652	4	US-09-427-770-24	Sequence 24, Appl
559	19	100.0	652	3	US-08-993-170A-26	Sequence 26, Appl	632	19	100.0	652	4	US-09-427-770-26	Sequence 26, Appl
560	19	100.0	652	3	US-08-993-170A-28	Sequence 28, Appl	633	19	100.0	652	4	US-09-427-770-28	Sequence 28, Appl
561	19	100.0	652	3	US-08-993-170A-30	Sequence 30, Appl	634	19	100.0	652	4	US-09-427-770-30	Sequence 30, Appl
562	19	100.0	652	3	US-08-993-170A-32	Sequence 32, Appl	635	19	100.0	652	4	US-09-427-770-32	Sequence 32, Appl
563	19	100.0	652	3	US-08-993-170A-34	Sequence 34, Appl	636	19	100.0	652	4	US-09-427-770-34	Sequence 34, Appl
564	19	100.0	652	3	US-08-993-170A-36	Sequence 36, Appl	637	19	100.0	652	4	US-09-427-770-36	Sequence 36, Appl
565	19	100.0	652	3	US-08-993-170A-38	Sequence 38, Appl	638	19	100.0	652	4	US-09-427-770-38	Sequence 38, Appl
566	19	100.0	652	3	US-08-993-170A-40	Sequence 40, Appl	639	19	100.0	652	4	US-09-427-770-40	Sequence 40, Appl
567	19	100.0	652	3	US-08-993-170A-42	Sequence 42, Appl	640	19	100.0	652	4	US-09-427-770-42	Sequence 42, Appl
568	19	100.0	652	3	US-08-993-170A-44	Sequence 44, Appl	641	19	100.0	652	4	US-09-427-770-44	Sequence 44, Appl
569	19	100.0	652	3	US-08-993-170A-46	Sequence 46, Appl	642	19	100.0	652	4	US-09-427-770-46	Sequence 46, Appl
570	19	100.0	652	3	US-08-993-170A-48	Sequence 48, Appl	643	19	100.0	652	4	US-09-427-770-48	Sequence 48, Appl
571	19	100.0	652	3	US-08-993-170A-50	Sequence 50, Appl	644	19	100.0	652	4	US-09-427-770-50	Sequence 50, Appl
572	19	100.0	652	3	US-08-993-170A-54	Sequence 54, Appl	645	19	100.0	652	4	US-09-427-770-54	Sequence 54, Appl
573	19	100.0	652	3	US-08-993-170A-60	Sequence 60, Appl	646	19	100.0	652	4	US-09-427-770-60	Sequence 60, Appl
574	19	100.0	652	3	US-08-993-170A-62	Sequence 62, Appl	647	19	100.0	652	4	US-09-427-770-62	Sequence 62, Appl
575	19	100.0	652	3	US-08-993-170A-64	Sequence 64, Appl	648	19	100.0	652	4	US-09-427-770-64	Sequence 64, Appl
576	19	100.0	652	3	US-08-993-170A-66	Sequence 66, Appl	649	19	100.0	652	4	US-09-427-770-66	Sequence 66, Appl
577	19	100.0	652	3	US-08-993-170A-68	Sequence 68, Appl	650	19	100.0	652	4	US-09-427-770-68	Sequence 68, Appl
578	19	100.0	652	3	US-08-993-170A-98	Sequence 98, Appl	651	19	100.0	652	4	US-09-427-770-98	Sequence 98, Appl
579	19	100.0	652	3	US-08-993-170A-108	Sequence 108, Appl	652	19	100.0	652	4	US-09-427-770-108	Sequence 108, Appl
580	19	100.0	652	3	US-08-993-170A-110	Sequence 110, Appl	653	19	100.0	652	4	US-09-427-770-110	Sequence 110, Appl
581	19	100.0	652	3	US-08-993-170A-111	Sequence 111, Appl	654	19	100.0	652	4	US-09-427-770-111	Sequence 111, Appl
582	19	100.0	652	3	US-08-993-775B-2	Sequence 2, Appl	655	19	100.0	652	4	US-09-427-769-2	Sequence 2, Appl
583	19	100.0	652	3	US-08-993-775B-4	Sequence 4, Appl	656	19	100.0	652	4	US-09-427-769-4	Sequence 4, Appl
584	19	100.0	652	3	US-08-993-775B-6	Sequence 6, Appl	657	19	100.0	652	4	US-09-427-769-6	Sequence 6, Appl
585	19	100.0	652	3	US-08-993-775B-8	Sequence 8, Appl	658	19	100.0	652	4	US-09-427-769-8	Sequence 8, Appl
586	19	100.0	652	3	US-08-993-775B-10	Sequence 10, Appl	659	19	100.0	652	4	US-09-427-769-10	Sequence 10, Appl
587	19	100.0	652	3	US-08-993-775B-12	Sequence 12, Appl	660	19	100.0	652	4	US-09-427-769-12	Sequence 12, Appl
588	19	100.0	652	3	US-08-993-775B-14	Sequence 14, Appl	661	19	100.0	652	4	US-09-427-769-14	Sequence 14, Appl
589	19	100.0	652	3	US-08-993-775B-16	Sequence 16, Appl	662	19	100.0	652	4	US-09-427-769-16	Sequence 16, Appl
590	19	100.0	652	3	US-08-993-775B-18	Sequence 18, Appl	663	19	100.0	652	4	US-09-427-769-18	Sequence 18, Appl
591	19	100.0	652	3	US-08-993-775B-20	Sequence 20, Appl	664	19	100.0	652	4	US-09-427-769-20	Sequence 20, Appl
592	19	100.0	652	3	US-08-993-775B-22	Sequence 22, Appl	665	19	100.0	652	4	US-09-427-769-22	Sequence 22, Appl
593	19	100.0	652	3	US-08-993-775B-24	Sequence 24, Appl	666	19	100.0	652	4	US-09-427-769-24	Sequence 24, Appl
594	19	100.0	652	3	US-08-993-775B-26	Sequence 26, Appl	667	19	100.0	652	4	US-09-427-769-26	Sequence 26, Appl
595	19	100.0	652	3	US-08-993-775B-28	Sequence 28, Appl	668	19	100.0	652	4	US-09-427-769-28	Sequence 28, Appl
596	19	100.0	652	3	US-08-993-775B-30	Sequence 30, Appl	669	19	100.0	652	4	US-09-427-769-30	Sequence 30, Appl
597	19	100.0	652	3	US-08-993-775B-32	Sequence 32, Appl	670	19	100.0	652	4	US-09-427-769-32	Sequence 32, Appl
598	19	100.0	652	3	US-08-993-775B-34	Sequence 34, Appl	671	19	100.0	652	4	US-09-427-769-34	Sequence 34, Appl
599	19	100.0	652	3	US-08-993-775B-36	Sequence 36, Appl	672	19	100.0	652	4	US-09-427-769-36	Sequence 36, Appl
600	19	100.0	652	3	US-08-993-775B-38	Sequence 38, Appl	673	19	100.0	652	4	US-09-427-769-38	Sequence 38, Appl
601	19	100.0	652	3	US-08-993-775B-40	Sequence 40, Appl	674	19	100.0	652	4	US-09-427-769-40	Sequence 40, Appl
602	19	100.0	652	3	US-08-993-775B-42	Sequence 42, Appl	675	19	100.0	652	4	US-09-427-769-42	Sequence 42, Appl
603	19	100.0	652	3	US-08-993-775B-44	Sequence 44, Appl	676	19	100.0	652	4	US-09-427-769-44	Sequence 44, Appl
604	19	100.0	652	3	US-08-993-775B-46	Sequence 46, Appl	677	19	100.0	652	4	US-09-427-769-46	Sequence 46, Appl
605	19	100.0	652	3	US-08-993-775B-48	Sequence 48, Appl	678	19	100.0	652	4	US-09-427-769-48	Sequence 48, Appl
606	19	100.0	652	3	US-08-993-775B-50	Sequence 50, Appl	679	19	100.0	652	4	US-09-427-769-50	Sequence 50, Appl
607	19	100.0	652	3	US-08-993-775B-54	Sequence 54, Appl	680	19	100.0	652	4	US-09-427-769-54	Sequence 54, Appl
608	19	100.0	652	3	US-08-993-775B-60	Sequence 60, Appl	681	19	100.0	652	4	US-09-427-769-60	Sequence 60, Appl
609	19	100.0	652	3	US-08-993-775B-62	Sequence 62, Appl	682	19	100.0	652	4	US-09-427-769-62	Sequence 62, Appl
610	19	100.0	652	3	US-08-993-775B-64	Sequence 64, Appl	683	19	100.0	652	4	US-09-427-769-64	Sequence 64, Appl
611	19	100.0	652	3	US-08-993-775B-66	Sequence 66, Appl	684	19	100.0	652	4	US-09-427-769-66	Sequence 66, Appl

685 19 100.0 652 4 US-09-427-769-68 Sequence 68, Appl  
686 19 100.0 652 4 US-09-427-769-98 Sequence 98, Appl  
687 19 100.0 652 4 US-09-427-769-108 Sequence 108, Appl  
688 19 100.0 652 4 US-09-427-769-110 Sequence 110, Appl  
689 19 100.0 652 4 US-09-427-769-111 Sequence 111, Appl  
690 19 100.0 652 5 PCT-US92-00040-2 Sequence 2, Appl  
691 19 100.0 652 6 5187091-2 Patent No. 5187091  
692 19 100.0 653 3 US-08-996-441B-100 Sequence 100, Appl  
693 19 100.0 653 3 US-08-993-722A-100 Sequence 100, Appl  
694 19 100.0 653 3 US-08-993-170A-100 Sequence 100, Appl  
695 19 100.0 653 3 US-08-993-775B-100 Sequence 100, Appl  
696 19 100.0 653 4 US-09-377-466B-8 Sequence 8, Appl  
697 19 100.0 653 4 US-09-377-466B-10 Sequence 10, Appl  
698 19 100.0 653 4 US-09-377-466B-12 Sequence 12, Appl  
699 19 100.0 653 4 US-09-377-466B-14 Sequence 14, Appl  
700 19 100.0 653 4 US-09-377-466B-16 Sequence 16, Appl  
701 19 100.0 653 4 US-09-377-466B-18 Sequence 18, Appl  
702 19 100.0 653 4 US-09-377-466B-20 Sequence 20, Appl  
703 19 100.0 653 4 US-09-377-466B-22 Sequence 22, Appl  
704 19 100.0 653 4 US-09-377-466B-24 Sequence 24, Appl  
705 19 100.0 653 4 US-09-377-466B-27 Sequence 27, Appl  
706 19 100.0 653 4 US-09-377-466B-39 Sequence 39, Appl  
707 19 100.0 653 4 US-09-427-770-100 Sequence 100, Appl  
708 19 100.0 653 4 US-09-427-770-100 Sequence 100, Appl  
709 19 100.0 654 1 US-08-487-890A-12 Sequence 12, Appl  
710 19 100.0 654 2 US-08-478-435-12 Sequence 12, Appl  
711 19 100.0 654 2 US-08-377-483-12 Sequence 12, Appl  
712 19 100.0 654 2 US-08-478-435-12 Sequence 12, Appl  
713 19 100.0 654 3 US-08-474-671-12 Sequence 12, Appl  
714 19 100.0 654 3 US-08-483-577A-12 Sequence 12, Appl  
715 19 100.0 654 3 US-08-897-438-12 Sequence 12, Appl  
716 19 100.0 654 4 US-08-637-654-12 Sequence 12, Appl  
717 19 100.0 654 4 US-08-649-518-12 Sequence 12, Appl  
718 19 100.0 657 3 US-08-613-009A-19 Sequence 19, Appl  
719 19 100.0 657 4 US-08-778-570B-25 Sequence 25, Appl  
720 19 100.0 657 4 US-09-059-584-25 Sequence 25, Appl  
721 19 100.0 657 4 US-09-252-991A-27358 Sequence 27358, A  
722 19 100.0 660 1 US-08-487-890A-8 Sequence 8, Appl  
723 19 100.0 660 1 US-08-487-890A-10 Sequence 8, Appl  
724 19 100.0 660 2 US-08-478-435-8 Sequence 8, Appl  
725 19 100.0 660 2 US-08-478-435-10 Sequence 8, Appl  
726 19 100.0 660 2 US-08-337-483-8 Sequence 8, Appl  
727 19 100.0 660 2 US-08-337-483-10 Sequence 8, Appl  
728 19 100.0 660 2 US-08-478-373-8 Sequence 8, Appl  
729 19 100.0 660 3 US-08-474-671-8 Sequence 8, Appl  
730 19 100.0 660 3 US-08-474-671-10 Sequence 8, Appl  
731 19 100.0 660 3 US-08-483-577A-8 Sequence 8, Appl  
732 19 100.0 660 3 US-08-483-577A-10 Sequence 8, Appl  
733 19 100.0 660 3 US-08-897-438-8 Sequence 8, Appl  
734 19 100.0 660 3 US-08-897-438-10 Sequence 8, Appl  
735 19 100.0 660 4 US-08-637-654-8 Sequence 8, Appl  
736 19 100.0 660 4 US-08-637-654-10 Sequence 8, Appl  
737 19 100.0 660 4 US-08-649-518-8 Sequence 8, Appl  
738 19 100.0 660 4 US-08-649-518-10 Sequence 8, Appl  
739 19 100.0 660 4 US-08-649-518-12 Sequence 8, Appl  
740 19 100.0 678 1 US-08-282-141-2 Sequence 2, Appl  
741 19 100.0 678 1 US-08-433-434-2 Sequence 2, Appl  
742 19 100.0 678 1 US-08-433-436-2 Sequence 2, Appl  
743 19 100.0 678 2 US-08-438-863-2 Sequence 2, Appl  
744 19 100.0 678 3 US-08-438-864-2 Sequence 2, Appl  
745 19 100.0 678 3 US-08-438-862-2 Sequence 2, Appl  
746 19 100.0 678 3 US-08-628-747-2 Sequence 2, Appl  
747 19 100.0 678 3 US-08-404-253-2 Sequence 2, Appl  
748 19 100.0 678 3 US-08-443-866B-2 Sequence 2, Appl  
749 19 100.0 680 4 US-09-252-991A-26639 Sequence 26639, A  
750 19 100.0 680 4 US-09-252-991A-24919 Sequence 24919, A  
751 19 100.0 685 4 US-09-252-991A-26723 Sequence 26723, A  
752 19 100.0 696 6 5262177-4 Patent No. 5262177  
753 19 100.0 698 2 US-08-175-158A-2 Sequence 2, Appl  
754 19 100.0 703 4 US-09-252-991A-23905 Sequence 23905, A  
755 19 100.0 718 4 US-09-252-991A-33109 Sequence 33109, A  
756 19 100.0 720 4 US-09-252-991A-18618 Sequence 18618, A  
757 19 100.0 720 4 US-09-134-000C-5631 Sequence 5631, Ap

758 19 100.0 736 4 US-09-252-991A-27827 Sequence 27827, A  
759 19 100.0 748 4 US-09-252-991A-28449 Sequence 28449, A  
760 19 100.0 749 4 US-09-562-737-100 Sequence 100, Appl  
761 19 100.0 753 4 US-09-252-991A-16877 Sequence 16877, A  
762 19 100.0 753 4 US-09-328-352-5412 Sequence 5412, Ap  
763 19 100.0 771 4 US-09-252-991A-17243 Sequence 17243, A  
764 19 100.0 788 2 US-08-907-166-6 Sequence 6, Appl  
765 19 100.0 788 4 US-09-391-340-6 Sequence 6, Appl  
766 19 100.0 789 4 US-09-390-234-16 Sequence 16, Appl  
767 19 100.0 789 4 US-09-603-311-16 Sequence 16, Appl  
768 19 100.0 791 4 US-09-107-532A-6065 Sequence 6065, Ap  
769 19 100.0 812 4 US-09-513-783A-4 Sequence 4, Appl  
770 19 100.0 812 4 US-09-513-783A-6 Sequence 6, Appl  
771 19 100.0 816 4 US-09-489-039A-10520 Sequence 10520, A  
772 19 100.0 816 4 US-09-651-656-15 Sequence 15, Appl  
773 19 100.0 819 4 US-09-650-855-15 Sequence 15, Appl  
774 19 100.0 821 3 US-09-422-869-24 Sequence 24, Appl  
775 19 100.0 844 3 US-09-029-267-20 Sequence 20, Appl  
776 19 100.0 846 4 US-09-328-352-6648 Sequence 6648, Ap  
777 19 100.0 848 4 US-09-489-039A-13773 Sequence 13773, A  
778 19 100.0 849 4 US-08-252-991A-17953 Sequence 17953, A  
779 19 100.0 925 4 US-09-252-991A-30504 Sequence 30504, A  
780 19 100.0 925 4 US-09-252-991A-27057 Sequence 27057, A  
781 19 100.0 927 4 US-09-134-001C-4831 Sequence 4831, Ap  
782 19 100.0 944 4 US-09-328-352-4401 Sequence 4401, Ap  
783 19 100.0 972 3 US-08-335-844A-24 Sequence 24, Appl  
784 19 100.0 972 4 US-09-129-366-24 Sequence 24, Appl  
785 19 100.0 975 4 US-08-543-681A-5755 Sequence 5755, Ap  
786 19 100.0 982 4 US-09-083-905A-7 Sequence 7, Appl  
787 19 100.0 990 4 US-09-252-991A-22792 Sequence 22792, A  
788 19 100.0 993 4 US-09-894-998A-50 Sequence 50, Appl  
789 19 100.0 1009 4 US-09-693-146-4 Sequence 4, Appl  
790 19 100.0 1037 4 US-09-894-998A-54 Sequence 54, Appl  
791 19 100.0 1056 4 US-09-513-783A-32 Sequence 32, Appl  
792 19 100.0 1061 4 US-08-252-991A-23691 Sequence 23691, A  
793 19 100.0 1074 2 US-08-470-058-2 Sequence 2, Appl  
794 19 100.0 1074 3 US-09-037-188-2 Sequence 2, Appl  
795 19 100.0 1074 3 US-09-285-310-2 Sequence 2, Appl  
796 19 100.0 1113 4 US-09-894-998A-51 Sequence 51, Appl  
797 19 100.0 1114 4 US-09-975-413A-12 Sequence 12, Appl  
798 19 100.0 1125 4 US-09-513-783A-152 Sequence 152, App  
799 19 100.0 1128 4 US-08-252-991A-19672 Sequence 19672, A  
800 19 100.0 1155 4 US-08-543-681A-6286 Sequence 6286, Ap  
801 19 100.0 1410 2 US-08-470-058-4 Sequence 4, Appl  
802 19 100.0 1410 3 US-09-037-188-4 Sequence 4, Appl  
803 19 100.0 1410 3 US-09-285-310-4 Sequence 4, Appl  
804 19 100.0 1417 3 US-08-452-459-10 Sequence 10, Appl  
805 19 100.0 1477 3 US-08-423-752-10 Sequence 10, Appl  
806 19 100.0 1477 3 US-08-945-994-3 Sequence 3, Appl  
807 19 100.0 1477 3 US-08-716-873-24 Sequence 24, Appl  
808 19 100.0 1477 3 US-09-368-431-24 Sequence 24, Appl  
809 19 100.0 1477 4 US-09-414-006-10 Sequence 10, Appl  
810 19 100.0 1477 4 US-09-447-223-10 Sequence 10, Appl  
811 19 100.0 1490 4 US-09-252-991A-28442 Sequence 28442, A  
812 19 100.0 1610 4 US-09-513-783A-22 Sequence 22, Appl  
813 19 100.0 2165 1 US-08-514-975B-2 Sequence 2, Appl  
814 19 100.0 2165 5 PCT-US95-12507-2 Sequence 2, Appl  
815 19 100.0 2227 3 US-08-475-886-2 Sequence 2, Appl  
816 19 100.0 2227 3 US-08-475-886-4 Sequence 4, Appl  
817 19 100.0 2227 3 US-08-475-886-6 Sequence 6, Appl  
818 19 100.0 2227 3 US-08-397-232-2 Sequence 2, Appl  
819 19 100.0 2227 3 US-08-397-232-4 Sequence 4, Appl  
820 19 100.0 2227 3 US-08-171-387-2 Sequence 2, Appl  
821 19 100.0 2227 4 US-09-653-499-2 Sequence 2, Appl  
822 19 100.0 2227 4 US-09-653-499-4 Sequence 4, Appl  
823 19 100.0 2227 4 US-09-653-499-6 Sequence 6, Appl  
824 19 100.0 2227 4 US-10-104-966-12 Sequence 12, Appl  
825 19 100.0 2227 4 US-10-135-988-2 Sequence 2, Appl  
826 19 100.0 2227 4 US-10-135-988-4 Sequence 4, Appl  
827 19 100.0 2227 4 US-10-135-988-6 Sequence 6, Appl  
828 19 100.0 2308 1 US-08-015-973-1 Sequence 1, Appl  
829 19 100.0 2308 2 US-08-448-164-1 Sequence 1, Appl  
830 19 100.0 2308 3 US-08-081-529-2 Sequence 2, Appl

831	19	100.0	2314	4	US-09-816-703A-2	Sequence 2, Appli	Sequence 2, Appli	18	94.7	149	1	US-08-100-874-2	Sequence 2, Appli
832	19	100.0	2629	4	US-08-751-189-4	Sequence 4, Appli	Sequence 4, Appli	18	94.7	149	1	US-08-963-409-4	Sequence 2, Appli
833	19	100.0	2629	2	US-09-060-836-4	Sequence 4, Appli	Sequence 4, Appli	18	94.7	149	3	US-08-641-873-20	Sequence 20, Appli
834	19	100.0	2629	3	US-09-184-444-4	Sequence 4, Appli	Sequence 4, Appli	18	94.7	152	3	US-08-963-409-5	Sequence 5, Appli
835	19	100.0	5588	3	US-09-036-987A-6	Sequence 6, Appli	Sequence 6, Appli	18	94.7	152	4	US-09-252-991A-18700	Sequence 18700, A
836	19	100.0	5588	3	US-09-370-700-6	Sequence 6, Appli	Sequence 6, Appli	18	94.7	152	4	US-09-252-991A-31418	Sequence 31418, A
837	19	100.0	5588	4	US-09-603-207-6	Sequence 6, Appli	Sequence 6, Appli	18	94.7	157	4	US-09-134-001C-2991	Sequence 2991, Ap
838	19	100.0	8991	4	US-08-714-741-32	Sequence 32, Appli	Sequence 32, Appli	18	94.7	158	1	US-08-611-107-4	Sequence 4, Appli
839	18	94.7	9	4	US-09-372-338-6	Sequence 6, Appli	Sequence 6, Appli	18	94.7	158	2	US-08-423-560A-4	Sequence 4, Appli
840	18	94.7	9	4	US-10-126-752-6	Sequence 6, Appli	Sequence 6, Appli	18	94.7	158	3	US-08-468-733-4	Sequence 4, Appli
841	18	94.7	10	2	US-08-482-228-120	Sequence 120, App	Sequence 120, App	18	94.7	160	2	US-08-602-941-1	Sequence 1, Appli
842	18	94.7	10	2	US-08-482-228-120	Sequence 120, App	Sequence 120, App	18	94.7	160	3	US-08-961-264-1	Sequence 1, Appli
843	18	94.7	11	2	US-08-934-223-99	Sequence 99, Appli	Sequence 99, Appli	18	94.7	160	4	US-09-442-039A-1	Sequence 1, Appli
844	18	94.7	11	2	US-08-933-402-99	Sequence 99, Appli	Sequence 99, Appli	18	94.7	160	4	US-09-612-342-1	Sequence 1, Appli
845	18	94.7	11	2	US-09-207-621-99	Sequence 99, Appli	Sequence 99, Appli	18	94.7	160	4	US-09-612-342-1	Sequence 1, Appli
846	18	94.7	11	2	US-08-532-818-99	Sequence 99, Appli	Sequence 99, Appli	18	94.7	160	4	US-09-724-401-1	Sequence 1, Appli
847	18	94.7	11	3	US-09-231-797-99	Sequence 99, Appli	Sequence 99, Appli	18	94.7	162	3	US-09-108-020-50	Sequence 50, Appli
848	18	94.7	11	3	US-08-934-224-99	Sequence 99, Appli	Sequence 99, Appli	18	94.7	172	4	US-09-107-532A-5847	Sequence 5847, Ap
849	18	94.7	11	3	US-08-933-843-99	Sequence 99, Appli	Sequence 99, Appli	18	94.7	172	4	US-09-634-238-396	Sequence 238, App
850	18	94.7	11	3	US-08-934-223-99	Sequence 99, Appli	Sequence 99, Appli	18	94.7	179	4	US-09-489-039A-11491	Sequence 11491, A
851	18	94.7	11	3	US-09-413-492-99	Sequence 99, Appli	Sequence 99, Appli	18	94.7	184	4	US-07-741-940-6	Sequence 6, Appli
852	18	94.7	19	4	US-08-835-281-10	Sequence 10, Appli	Sequence 10, Appli	18	94.7	185	1	US-08-289-548A-6	Sequence 6, Appli
853	18	94.7	20	1	US-08-658-130-1	Sequence 1, Appli	Sequence 1, Appli	18	94.7	185	1	US-08-453-654-6	Sequence 6, Appli
854	18	94.7	20	2	US-08-799-825-22	Sequence 22, Appli	Sequence 22, Appli	18	94.7	185	1	US-08-452-655B-6	Sequence 6, Appli
855	18	94.7	20	4	US-09-092-000-5	Sequence 5, Appli	Sequence 5, Appli	18	94.7	185	2	US-08-863-336-3	Sequence 3, Appli
856	18	94.7	20	4	US-09-280-068-22	Sequence 22, Appli	Sequence 22, Appli	18	94.7	185	2	US-08-863-336-3	Sequence 3, Appli
857	18	94.7	20	4	US-08-835-281-12	Sequence 12, Appli	Sequence 12, Appli	18	94.7	185	3	US-08-450-582-6	Sequence 6, Appli
858	18	94.7	20	4	US-09-481-593-18	Sequence 18, Appli	Sequence 18, Appli	18	94.7	185	4	US-08-449-731-6	Sequence 6, Appli
859	18	94.7	20	5	PCT-US96-08723-1	Sequence 1, Appli	Sequence 1, Appli	18	94.7	185	4	US-09-198-452A-858	Sequence 858, App
860	18	94.7	22	2	US-08-907-673-1	Sequence 6, Appli	Sequence 6, Appli	18	94.7	187	4	US-09-134-001C-3419	Sequence 3419, Ap
861	18	94.7	22	3	US-08-809-397-6	Sequence 6, Appli	Sequence 6, Appli	18	94.7	189	4	US-09-134-000C-4845	Sequence 4845, Ap
862	18	94.7	22	5	PCT-US95-12502-6	Sequence 6, Appli	Sequence 6, Appli	18	94.7	199	3	US-09-602-282-2	Sequence 2, Appli
863	18	94.7	22	5	PCT-US95-12502-6	Sequence 6, Appli	Sequence 6, Appli	18	94.7	199	3	US-09-004-716-2	Sequence 2, Appli
864	18	94.7	27	4	US-09-039-780A-82	Sequence 82, Appli	Sequence 82, Appli	18	94.7	199	4	US-09-477-510-2	Sequence 2, Appli
865	18	94.7	28	4	US-09-039-780A-81	Sequence 81, Appli	Sequence 81, Appli	18	94.7	199	4	US-09-543-681A-5219	Sequence 5219, Ap
866	18	94.7	30	4	US-09-039-780A-104	Sequence 104, App	Sequence 104, App	18	94.7	199	5	PCT-US94-02889-2	Sequence 2, Appli
867	18	94.7	38	4	US-09-039-780A-93	Sequence 83, Appli	Sequence 83, Appli	18	94.7	201	1	US-07-928-580B-5	Sequence 5, Appli
868	18	94.7	38	4	US-09-039-780A-84	Sequence 84, Appli	Sequence 84, Appli	18	94.7	205	4	US-09-107-532A-5214	Sequence 5214, Ap
869	18	94.7	39	2	US-08-889-291-15	Sequence 15, Appli	Sequence 15, Appli	18	94.7	208	1	US-09-631-607-3	Sequence 3, Appli
870	18	94.7	39	3	US-09-098-244-15	Sequence 15, Appli	Sequence 15, Appli	18	94.7	208	4	US-09-098-358B-3	Sequence 3, Appli
871	18	94.7	39	4	US-09-375-314-15	Sequence 15, Appli	Sequence 15, Appli	18	94.7	210	1	US-07-741-940-4	Sequence 4, Appli
872	18	94.7	39	4	US-09-767-395-15	Sequence 15, Appli	Sequence 15, Appli	18	94.7	210	1	US-08-289-548A-4	Sequence 4, Appli
873	18	94.7	43	4	US-09-039-780A-89	Sequence 89, Appli	Sequence 89, Appli	18	94.7	210	1	US-08-452-655B-4	Sequence 4, Appli
874	18	94.7	44	4	US-09-039-780A-98	Sequence 98, Appli	Sequence 98, Appli	18	94.7	210	1	US-08-452-655B-4	Sequence 4, Appli
875	18	94.7	44	4	US-09-039-780A-99	Sequence 99, Appli	Sequence 99, Appli	18	94.7	210	3	US-08-450-582-4	Sequence 4, Appli
876	18	94.7	54	4	US-09-039-780A-90	Sequence 90, Appli	Sequence 90, Appli	18	94.7	210	4	US-08-448-731-4	Sequence 4, Appli
877	18	94.7	56	4	US-09-039-780A-93	Sequence 93, Appli	Sequence 93, Appli	18	94.7	211	4	US-09-328-352-5256	Sequence 5256, Ap
878	18	94.7	60	4	US-09-134-000C-4258	Sequence 4258, Ap	Sequence 4258, Ap	18	94.7	214	4	US-09-795-926-16	Sequence 16, Appli
879	18	94.7	61	4	US-09-107-532A-4993	Sequence 4993, Ap	Sequence 4993, Ap	18	94.7	220	4	US-09-107-532A-6803	Sequence 6803, Ap
880	18	94.7	66	4	US-09-543-681A-5898	Sequence 5898, Ap	Sequence 5898, Ap	18	94.7	223	1	US-07-708-885B-3	Sequence 3, Appli
881	18	94.7	69	4	US-09-328-352-5719	Sequence 5719, Ap	Sequence 5719, Ap	18	94.7	223	1	US-07-714-386-3	Sequence 3, Appli
882	18	94.7	88	4	US-09-732-210-1423	Sequence 1423, Ap	Sequence 1423, Ap	18	94.7	223	1	US-07-708-888A-3	Sequence 3, Appli
883	18	94.7	90	4	US-09-393-634-76	Sequence 76, Appli	Sequence 76, Appli	18	94.7	229	4	US-09-483-039A-9391	Sequence 9391, Ap
884	18	94.7	101	4	US-09-107-532A-6901	Sequence 6901, Ap	Sequence 6901, Ap	18	94.7	232	4	US-09-134-000C-3556	Sequence 3556, Ap
885	18	94.7	105	4	US-09-489-039A-10362	Sequence 10362, A	Sequence 10362, A	18	94.7	232	4	US-09-390-721-4	Sequence 4, Appli
886	18	94.7	113	4	US-09-614-912-186	Sequence 186, App	Sequence 186, App	18	94.7	233	3	US-09-248-335-48	Sequence 48, Appli
887	18	94.7	116	4	US-09-489-039A-12512	Sequence 12512, A	Sequence 12512, A	18	94.7	243	4	US-09-134-000C-3753	Sequence 3753, Ap
888	18	94.7	116	4	US-09-540-236-3531	Sequence 3531, Ap	Sequence 3531, Ap	18	94.7	245	4	US-09-543-681A-5699	Sequence 5699, Ap
889	18	94.7	122	4	US-09-540-236-3530	Sequence 3530, Ap	Sequence 3530, Ap	18	94.7	247	4	US-09-328-352-7311	Sequence 7311, Ap
890	18	94.7	131	4	US-09-328-352-4806	Sequence 4806, Ap	Sequence 4806, Ap	18	94.7	249	4	US-09-252-991A-32233	Sequence 32233, A
891	18	94.7	132	4	US-09-134-000C-3642	Sequence 3642, Ap	Sequence 3642, Ap	18	94.7	251	4	US-09-648-004-20	Sequence 20, Appli
892	18	94.7	136	4	US-09-732-210-1423	Sequence 211, App	Sequence 211, App	18	94.7	251	4	US-09-286-981B-4	Sequence 4, Appli
893	18	94.7	142	1	US-07-951-715A-24	Sequence 24, Appli	Sequence 24, Appli	18	94.7	252	4	US-09-489-039A-13690	Sequence 13690, A
894	18	94.7	142	2	US-08-459-448A-24	Sequence 24, Appli	Sequence 24, Appli	18	94.7	257	2	US-09-198-452A-632	Sequence 632, App
895	18	94.7	142	3	US-08-459-504B-24	Sequence 24, Appli	Sequence 24, Appli	18	94.7	257	2	US-08-637-759B-92	Sequence 92, Appli
896	18	94.7	142	3	US-08-459-444A-24	Sequence 24, Appli	Sequence 24, Appli	18	94.7	257	3	US-08-871-355A-92	Sequence 92, Appli
897	18	94.7	142	3	US-08-459-444A-24	Sequence 24, Appli	Sequence 24, Appli	18	94.7	257	3	US-08-621-353-9	Sequence 9, Appli
898	18	94.7	142	4	US-09-547-423-24	Sequence 24, Appli	Sequence 24, Appli	18	94.7	262	1	US-08-622-352A-11	Sequence 11, Appli
899	18	94.7	143	3	US-09-328-352-7008	Sequence 7008, Ap	Sequence 7008, Ap	18	94.7	262	3	US-08-826-390-11	Sequence 11, Appli
900	18	94.7	143	4	US-09-328-352-7008	Sequence 7008, Ap	Sequence 7008, Ap	18	94.7	262	3	US-09-795-926-6	Sequence 6, Appli
901	18	94.7	143	4	US-09-134-000C-4544	Sequence 4544, Ap	Sequence 4544, Ap	18	94.7	265	4	US-09-134-000C-5370	Sequence 5370, Ap
902	18	94.7	144	4	US-09-543-681A-5427	Sequence 5427, Ap	Sequence 5427, Ap	18	94.7	265	4	US-09-543-681A-6389	Sequence 6389, Ap
903	18	94.7	145	3	US-08-720-625-5	Sequence 5, Appli	Sequence 5, Appli	18	94.7	267	4		

977 18 94.7 268 4 US-09-489-039A-7310 Sequence 7310, Ap  
 978 18 94.7 269 4 US-09-540-236-2149 Sequence 2149, Ap  
 979 18 94.7 270 4 US-09-489-039A-9313 Sequence 9313, Ap  
 980 18 94.7 273 4 US-09-489-039A-12815 Sequence 12815, A  
 981 18 94.7 280 4 US-09-252-991A-23529 Sequence 23529, A  
 982 18 94.7 281 3 US-09-053-702-2 Sequence 2, Appl  
 983 18 94.7 281 4 US-09-540-236-2349 Sequence 2349, Ap  
 984 18 94.7 282 4 US-09-372-422A-8 Sequence 8, Appl  
 985 18 94.7 284 4 US-09-372-448A-4 Sequence 4, Appl  
 986 18 94.7 285 4 US-09-314-701-30 Sequence 30, Appl  
 987 18 94.7 286 4 US-09-489-039A-13264 Sequence 13264, A  
 988 18 94.7 288 4 US-09-372-422A-12 Sequence 12, Appl  
 989 18 94.7 288 4 US-09-372-422A-16 Sequence 16, Appl  
 990 18 94.7 288 4 US-09-372-422A-18 Sequence 18, Appl  
 991 18 94.7 288 4 US-09-372-448A-2 Sequence 2, Appl  
 992 18 94.7 289 4 US-09-372-422A-2 Sequence 2, Appl  
 993 18 94.7 289 4 US-09-372-422A-14 Sequence 14, Appl  
 994 18 94.7 292 4 US-09-372-422A-4 Sequence 4, Appl  
 995 18 94.7 292 4 US-09-372-422A-10 Sequence 10, Appl  
 996 18 94.7 292 4 US-09-489-039A-9431 Sequence 9431, Ap  
 997 18 94.7 295 4 US-09-134-000C-3576 Sequence 3576, Ap  
 998 18 94.7 296 4 US-09-372-422A-20 Sequence 20, Appl  
 999 18 94.7 296 4 US-09-328-352-6427 Sequence 6427, Ap  
 1000 18 94.7 296 4 US-09-540-236-2843 Sequence 2843, Ap

## ALIGNMENTS

## RESULT 1

US-09-724-566A-78 ; Sequence 78, Application US/09724566A  
 ; Patent No. 6627739

; GENERAL INFORMATION:  
 ; APPLICANT: Anderson, John P.

; APPLICANT: Basi, Gurigbal  
 ; APPLICANT: Doane, Minh Tam

; APPLICANT: Prignon, No. 6627739mand  
 ; APPLICANT: John, Varghese

; APPLICANT: Power, Michael  
 ; APPLICANT: Sinha, Sukanto

; APPLICANT: Tatsuno, Gwen  
 ; APPLICANT: Tung, Jay

; APPLICANT: Wang, Shuwen  
 ; APPLICANT: McConlogue, Lisa

; TITLE OF INVENTION: Beta-Secretase Enzyme Compositions and  
 ; FILE REFERENCE: 228-US-NEWC2

; CURRENT APPLICATION NUMBER: US/09/724,566A  
 ; CURRENT FILING DATE: 2000-11-28

; PRIOR APPLICATION NUMBER: US 09/501,708  
 ; PRIOR FILING DATE: 2000-02-10

; PRIOR APPLICATION NUMBER: 60/119,571  
 ; PRIOR FILING DATE: 1999-02-10

; PRIOR APPLICATION NUMBER: 60/139,172  
 ; PRIOR FILING DATE: 1999-06-15

; NUMBER OF SEQ ID NOS: 104  
 ; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 78  
 ; LENGTH: 7

; TYPE: PRT  
 ; ORGANISM: Artificial Sequence

; FEATURE:  
 ; OTHER INFORMATION: Peptide inhibitor P3-P4' XD-V

; FEATURE:  
 ; NAME/KEY: MOD\_RES

; LOCATION: 3  
 ; OTHER INFORMATION: Xaa is hydroxyethylene or statine

US-09-724-566A-78

Query Match 100.0%; Score 19; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 |||||

Db 4 VAEF 7  
 |||||

## RESULT 2

US-09-724-566A-81

; Sequence 81, Application US/09724566A

; Patent No. 6627739

; GENERAL INFORMATION:  
 ; APPLICANT: Anderson, John P.

; APPLICANT: Basi, Gurigbal  
 ; APPLICANT: Doane, Minh Tam

; APPLICANT: Frignon, No. 6627739mand  
 ; APPLICANT: John, Varghese

; APPLICANT: Power, Michael  
 ; APPLICANT: Sinha, Sukanto

; APPLICANT: Tatsuno, Gwen  
 ; APPLICANT: Tung, Jay

; APPLICANT: Wang, Shuwen  
 ; APPLICANT: McConlogue, Lisa

; TITLE OF INVENTION: Beta-Secretase Enzyme Compositions and  
 ; FILE REFERENCE: 228-US-NEWC2

; CURRENT APPLICATION NUMBER: US/09/724,566A  
 ; CURRENT FILING DATE: 2000-11-28

; PRIOR APPLICATION NUMBER: US 09/501,708  
 ; PRIOR FILING DATE: 2000-02-10

; PRIOR APPLICATION NUMBER: 60/119,571  
 ; PRIOR FILING DATE: 1999-02-10

; PRIOR APPLICATION NUMBER: 60/139,172  
 ; PRIOR FILING DATE: 1999-06-15

; NUMBER OF SEQ ID NOS: 104  
 ; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 81  
 ; LENGTH: 8

; TYPE: PRT  
 ; ORGANISM: Artificial Sequence

; FEATURE:  
 ; OTHER INFORMATION: Peptide inhibitor P4-P4' XD-V

; FEATURE:  
 ; NAME/KEY: MOD\_RES

; LOCATION: 4  
 ; OTHER INFORMATION: Xaa is hydroxyethylene or statine

US-09-724-566A-81

Query Match 100.0%; Score 19; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 |||||

Db 5 VAEF 8  
 |||||

## RESULT 3

US-08-197-484-86

; Sequence 86, Application US/08197484

; Patent No. 6419931

; GENERAL INFORMATION:  
 ; APPLICANT: VITIELLO, Maria A.

; APPLICANT: CHESTNUT, Robert W.  
 ; APPLICANT: SETTE, Alessandro D.

; APPLICANT: CELIS, Esteban  
 ; APPLICANT: GRAY, Howard

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING  
 ; TITLE OF INVENTION: CTL IMMUNITY

; NUMBER OF SEQUENCES: 153  
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend Knourie and Crew  
 STREET: Steuart Street Tower, One Market Plaza

CITY: San Francisco

STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US 07/935,811  
FILING DATE: 26-AUG-1992  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/197,484  
FILING DATE: 16-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/935,811  
FILING DATE: 26-AUG-1992  
APPLICATION DATA:  
FILING DATE: 27-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/874,491  
FILING DATE: 29-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,682  
FILING DATE: 29-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/749,568  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Parmelee, Steven W.  
REGISTRATION NUMBER: 31,990  
REFERENCE/DOCKET NUMBER: 14137-26-4  
TELEPHONE: (206) 467-9600  
TELEFAX: (206) 623-6793  
INFORMATION FOR SEQ ID NO: 86:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
S-08-197-484-86

Query Match 100.0%; Score 19; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 VAEF 4  
b 2 VAEF 5

RESULT 4  
S-08-197-484-145  
Sequence 145, Application US/08197484  
Patent No. 6419931  
GENERAL INFORMATION:  
APPLICANT: VITIELLO, Maria A.  
APPLICANT: CHESTNUT, Robert W.  
APPLICANT: SETTE, Alessandro D.  
APPLICANT: CELIS, Esteban  
APPLICANT: GRAY, Howard  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING  
TITLE OF INVENTION: CTL IMMUNITY  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: Stewart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/197,484  
FILING DATE: 16-FEB-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/935,811  
FILING DATE: 26-AUG-1992  
APPLICATION DATA:  
FILING DATE: 27-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/874,491  
FILING DATE: 29-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,682  
FILING DATE: 29-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/749,568  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Parmelee, Steven W.  
REGISTRATION NUMBER: 31,990  
REFERENCE/DOCKET NUMBER: 14137-26-4  
TELEPHONE: (206) 467-9600  
TELEFAX: (206) 623-6793  
INFORMATION FOR SEQ ID NO: 145:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-197-484-145

Query Match 100.0%; Score 19; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 2 VAEF 5

RESULT 5  
US-09-724-566A-73  
Sequence 73, Application US/09724566A  
Patent No. 6627739  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Basi, Gurqbal  
APPLICANT: Doane, Minh Tam  
APPLICANT: Frigon, No. 6627739mand  
APPLICANT: John, Varghese  
APPLICANT: Power, Michael  
APPLICANT: Sinha, Sukanto  
APPLICANT: Tatsuno, Gwen  
APPLICANT: Tung, Jay  
APPLICANT: Wang, Shuwen  
APPLICANT: McCollough, Lisa  
TITLE OF INVENTION: Beta-Secretase Enzyme Compositions and  
TITLE OF INVENTION: Methods  
FILE REFERENCE: 228-US-NEWC2  
CURRENT APPLICATION NUMBER: US/09/724,566A  
CURRENT FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: US 09/501,708  
PRIOR FILING DATE: 2000-02-10  
PRIOR APPLICATION NUMBER: 60/119,571  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 60/139,172  
PRIOR FILING DATE: 1999-06-15  
NUMBER OF SEQ ID NOS: 104  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 73  
LENGTH: 9

TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: P4-P4'sta-d-V peptide inhibitor  
NAME/KEY: MOD\_RES  
LOCATION: 5  
OTHER INFORMATION: Xaa is statine moiety  
JS-09-724-566A-73

Query Match 100.0%; Score 19; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 6 VAEF 9

RESULT 6  
PCT-US95-02121-86  
; Sequence 86, Application PC/TUS9502121  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING  
; NUMBER OF SEQUENCES: 153  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/02121  
; FILING DATE: 16-FEB-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/197,484  
; FILING DATE: 16-FEB-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/935,811  
; FILING DATE: 26-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/874,491  
; FILING DATE: 27-APR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/827,682  
; FILING DATE: 29-JAN-1992  
; APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/02121  
; FILING DATE: 16-FEB-1995  
; CLASSIFICATION:  
; NAME: Parmelee, Steven W.  
; REGISTRATION NUMBER: 31,990  
; REFERENCE/DOCKET NUMBER: 14137-26-4PC  
; TELEPHONE: (206) 467-9600  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 86:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
PCT-US95-02121-86

Query Match 100.0%; Score 19; DB 5; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 6 VAEF 9

Db 2 VAEF 5

RESULT 7  
PCT-US95-02121-145  
; Sequence 145, Application PC/TUS9502121  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING  
; NUMBER OF SEQUENCES: 153  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/02121  
; FILING DATE: 16-FEB-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/197,484  
; FILING DATE: 16-FEB-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/935,811  
; FILING DATE: 26-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/874,491  
; FILING DATE: 27-APR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/827,682  
; FILING DATE: 29-JAN-1992  
; APPLICATION DATA:  
; APPLICATION NUMBER: US 07/749,568  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parmelee, Steven W.  
; REGISTRATION NUMBER: 31,990  
; REFERENCE/DOCKET NUMBER: 14137-26-4PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 467-9600  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 145:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: Peptide  
PCT-US95-02121-145

Query Match 100.0%; Score 19; DB 5; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 2 VAEF 5

RESULT 8  
US-09-724-566A-72  
; Sequence 72, Application US/09724566A  
; Patent No. 6627739  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, John P.  
; APPLICANT: Basi, Gurigbal  
; APPLICANT: Doane, Minh Tam  
; APPLICANT: Frigon, No. 6627739mand  
; APPLICANT: John, Varghese  
; APPLICANT: Power, Michael  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Tatsuno, Gwen

APPLICANT: Tung, Jay  
APPLICANT: Wang, Shuwen  
APPLICANT: McConlogue, Lisa  
TITLE OF INVENTION: Beta-Secretase Enzyme Compositions and  
METHODS  
FILE REFERENCE: 228-US-NEWC2  
CURRENT APPLICATION NUMBER: US/09/724,566A  
PRIOR FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: US 09/501,708  
PRIOR FILING DATE: 2000-02-10  
PRIOR APPLICATION NUMBER: 60/119,571  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 60/139,172  
PRIOR FILING DATE: 1999-06-15  
NUMBER OF SEQ ID NOS: 104  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 72  
LENGTH: 14  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: P10-P4' stad-v peptide inhibitor  
NAME/KEY: MOD\_RES  
LOCATION: 10  
OTHER INFORMATION: Xaa is statine moiety  
3-09-724-566A-72

Query Match 100.0%; Score 19; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 11 VAEF 14

RESULT 9  
S-09-724-566A-97  
Sequence 97, Application US/09724566A  
Patent No. 6627739  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Basi, Gurigbal  
APPLICANT: Doane, Minh Tam  
APPLICANT: Frigon, No. 6627739mand  
APPLICANT: John, Varghese  
APPLICANT: Power, Michael  
APPLICANT: Sinha, Sukanto  
APPLICANT: Tatsuno, Gwen  
APPLICANT: Tung, Jay  
APPLICANT: Wang, Shuwen  
APPLICANT: McConlogue, Lisa  
TITLE OF INVENTION: Beta-Secretase Enzyme Compositions and  
METHODS  
FILE REFERENCE: 228-US-NEWC2  
CURRENT APPLICATION NUMBER: US/09/724,566A  
CURRENT FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: US 09/501,708  
PRIOR FILING DATE: 2000-02-10  
PRIOR APPLICATION NUMBER: 60/119,571  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 60/139,172  
PRIOR FILING DATE: 1999-06-15  
NUMBER OF SEQ ID NOS: 104  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 97  
LENGTH: 14  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: APP-derived fragment P10-P4' (D-V)  
IS-09-724-566A-97

Query Match 100.0%; Score 19; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
DB 11 VAEF 14

RESULT 10  
US-08-733-825-3  
Sequence 3, Application US/08733825  
Patent No. 5837839  
GENERAL INFORMATION:  
APPLICANT: Toth, Matthew J.  
APPLICANT: Huwyler, Leslie R.  
TITLE OF INVENTION: Coding Sequences for Mevalonate  
TITLE OF INVENTION: Pyrophosphate Decarboxylase  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5837839artis Patent and Trademark Department  
STREET: 59 Route 10  
CITY: East Hanover  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07936-1080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/733,825  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/005,652  
FILING DATE: 18-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5837839ak, Henry P.  
REGISTRATION NUMBER: 33200  
REFERENCE/DOCKET NUMBER: 4-20615/Pl/CGC 1834  
TELEPHONE: (908) 277-5110  
TELEFAX: (908) 277-4606  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-733-825-3

Query Match 100.0%; Score 19; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 20 VAEF 23

RESULT 11  
US-08-487-890A-69  
Sequence 69, Application US/08487890A  
Patent No. 5708149  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele

APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Mardin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1S7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,890A  
FILING DATE: 07-JUN-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-487-890A-69

Query Match 100.0%; Score 19; DB 1; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 14 VAEF 17

RESULT 12  
US-08-468-763-1  
Sequence 1, Application US/08468763  
Patent No. 5741671  
GENERAL INFORMATION:  
APPLICANT: Agre, Peter C.  
TITLE OF INVENTION: Isolation, Cloning and Expression of  
TITLE OF INVENTION: Transmembrane Water Channel Proteins  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Allegretti  
STREET: 1001 G Street, N.W.  
CITY: Washington, D.C.  
STATE: D.C.  
COUNTRY: US  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,763  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/393,996  
FILING DATE: 24-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Posorske, Laurence H.  
REGISTRATION NUMBER: 34,698  
REFERENCE/DOCKET NUMBER: 1107.48633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 508-9100  
TELEFAX: 202 508-9299  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-468-763-1

Query Match 100.0%; Score 19; DB 1; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 14 VAEF 17

RESULT 13  
US-08-393-996A-1  
Sequence 1, Application US/08393996A  
Patent No. 5858702  
GENERAL INFORMATION:  
APPLICANT: Agre, Peter C.  
TITLE OF INVENTION: Isolation, Cloning and Expression of  
TITLE OF INVENTION: Transmembrane Water Channel Proteins  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Allegretti  
STREET: 1001 G Street, N.W.  
CITY: Washington, D.C.  
STATE: D.C.  
COUNTRY: US  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/393,996A  
FILING DATE: 24-FEB-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Posorske, Laurence H.  
REGISTRATION NUMBER: 34,698  
REFERENCE/DOCKET NUMBER: 1107.48633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 508-9100  
TELEFAX: 202 508-9299  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-393-996A-1

Query Match 100.0%; Score 19; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
|||  
b 14 VAEF 17

RESULT 14  
S-08-478-435-69  
Sequence 69, Application US/08478435  
Patent No. 5923323  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Murdin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,435  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/337,483  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-462 MIS.vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

S-08-478-435-69

Query Match 100.0%; Score 19; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
|||  
b 14 VAEF 17

RESULT 15  
US-08-337-483-69  
Sequence 69, Application US/08337483  
Patent No. 5922562  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Murdin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/337,483  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-410 MIS.jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-337-483-69

Query Match 100.0%; Score 19; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAEF 4  
|||  
Db 14 VAEF 17

RESULT 16  
US-08-478-373-69  
Sequence 69, Application US/08478373  
Patent No. 5922841  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Murdin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney

US-08-478-373-69

```

; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: MSG IR7
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,373
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-478-373-69
;
; Query Match 100.0%; Score 19; DB 2; Length 35;
; Best Local Similarity 100.0%; Pred. No. 1.3e+02;
; Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 VAEF 4
; DB 14 VAEF 17
;
; RESULT 17
; US-08-474-671-69
; Sequence 69, Application US/08474671
; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: MSG IR7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-474-671-69
;
; Query Match 100.0%; Score 19; DB 3; Length 35;
; Best Local Similarity 100.0%; Pred. No. 1.3e+02;
; Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 VAEF 4
; DB 14 VAEF 17
;
; RESULT 18
; US-08-483-577A-69
; Sequence 69, Application US/08483577A
; Patent No. 6015688
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: MSG IR7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,577A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483

```

FILING DATE: 08-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-511  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
S-08-483-577A-69

Query Match 100.0%; Score 19; DB 3; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
b 14 VAEF 17

RESULT 19  
S-08-897-438-69  
Sequence 69, Application US/08897438  
Patent No. 6262016  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Murdin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 160  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/897,438  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/483,577  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA: US 08/337,483  
APPLICATION NUMBER: US 08/337,483  
FILING DATE: 08-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968

FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-720  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-897-438-69

Query Match 100.0%; Score 19; DB 3; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 14 VAEF 17

RESULT 20  
US-08-637-654-69  
Sequence 69, Application US/08637654  
Patent No. 6358727  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Harkness, Robin E  
APPLICANT: Schryvers, Anthony B  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Murdin, Andrew D  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: HAEMOPHILUS TRANSFERRIN RECEPTOR GENES  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/637,654  
FILING DATE: 05-AUG-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA94/00616  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-595  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-637-654-69

Query Match 100.0%; Score 19; DB 4; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 14 VAEF 17

## RESULT 21

US-08-649-518-69  
; Sequence 69, Application US/08649518  
; Patent No. 6361779  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena  
; APPLICANT: Harkness, Robin  
; APPLICANT: Schryvers, Anthony  
; APPLICANT: Chong, Pele  
; APPLICANT: Gray-Owen, Scott  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Mordin, Andrew  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Transferrin Receptor Genes  
; NUMBER OF SEQUENCES: 160  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,518  
FILING DATE: 17-MAY-1996

CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/337,483  
FILING DATE: 08-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/483,577  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-608  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-649-518-69

Query Match 100.0%; Score 19; DB 4; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4

DB 14 VAEF 17

## RESULT 22

US-09-393-634-80  
; Sequence 80, Application US/09393634  
; Patent No. 6559910  
; GENERAL INFORMATION:  
; APPLICANT: Zuker, Charles S.  
; APPLICANT: Adler, Jon Elliot  
; APPLICANT: Ryba, Nick  
; APPLICANT: Mueller, Ken  
; APPLICANT: Hoon, Mark  
; APPLICANT: The Regents of the University of California  
; APPLICANT: The Government of the United States of America  
; APPLICANT: as represented by the Secretary of the  
; APPLICANT: Department of Health and Human Services  
; TITLE OF INVENTION: SF, a No. 6558910el Family of Taste Receptors  
; FILE REFERENCE: 02307E-098000US  
; CURRENT APPLICATION NUMBER: US/09/393,634  
; CURRENT FILING DATE: 1999-09-10  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 80  
; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human GR24  
US-09-393-634-80

Query Match 100.0%; Score 19; DB 4; Length 68;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 14 VAEF 17

## RESULT 23

US-09-489-039A-13555  
; Sequence 13555, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 13555  
; LENGTH: 69  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-13555

Query Match 100.0%; Score 19; DB 4; Length 69;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 5 VAEF 8

## RESULT 24

US-09-328-352-7111  
; Sequence 7111, Application US/09328352

Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 7111  
LENGTH: 71  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-7111

Query Match 100.0%; Score 19; DB 4; Length 71;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
|  
|  
|  
|  
26 VAEF 29

RESULT 25  
US-08-321-071A-10  
Sequence 10, Application US/08321071A  
Patent No. 5672686  
GENERAL INFORMATION:  
APPLICANT: CHITTENDEN, THOMAS D.  
TITLE OF INVENTION: APOPTOSIS RELATED PROTEIN Bcl-Y, AND METHODS  
FILE REFERENCE: OF USE THEREOF  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hale and Dorr  
STREET: 1455 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/321,071A  
FILING DATE: 11-OCT-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION NUMBER: PCT/US95/10103  
FILING DATE: 09-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/287,427  
FILING DATE: 09-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: WIXON, HENRY N.  
REGISTRATION NUMBER: 32,073  
REFERENCE/DOCKET NUMBER: 104322.121CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-942-8400  
TELEFAX: 202-942-8494  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 73 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: peptide  
US-08-321-071A-10

Query Match 100.0%; Score 19; DB 1; Length 73;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
|  
|  
|  
|  
26 VAEF 29

RESULT 26  
US-08-894-139-10  
Sequence 10, Application US/08894139  
Patent No. 6448376  
GENERAL INFORMATION:  
APPLICANT: LA THANGUE, NICHOLAS B.  
APPLICANT: BERNARDS, RENE  
APPLICANT: HIJMAN, ELEANORE M.  
TITLE OF INVENTION: TRANSCRIPTION FACTOR E2F-5  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHVE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,139  
FILING DATE: 13-AUG-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: WILSON, MARY J.  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 620-22  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 74 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-894-139-10

Query Match 100.0%; Score 19; DB 4; Length 74;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
|  
|  
|  
|  
23 VAEF 26

RESULT 27  
US-09-732-210-1412  
Sequence 1412, Application US/09732210  
Patent No. 6573361  
GENERAL INFORMATION:  
APPLICANT: Bunkers, Greg J.  
APPLICANT: Liang, Jihong  
APPLICANT: Mittanck, Cindy A.  
APPLICANT: Seale, Jeffrey W.  
APPLICANT: Wu, Yonnie S.  
TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use  
FILE REFERENCE: 38-21(15036)B  
CURRENT APPLICATION NUMBER: US/09/732,210  
CURRENT FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/169,513  
PRIOR FILING DATE: 1999-12-07

; PRIOR APPLICATION NUMBER: US 60/169,340  
; PRIOR FILING DATE: 1999-12-07  
; NUMBER OF SEQ ID NOS: 1753  
; SEQ ID NO 1412  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-732-210-1412

Query Match 100.0%; Score 19; DB 4; Length 88;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 11 VAEF 14

RESULT 28  
US-07-612-674-12  
; Sequence 12, Application US/07612674  
; Patent No. 5658792  
; GENERAL INFORMATION:  
; APPLICANT: NUEL, MARK J.  
; APPLICANT: MCCLUNG, J. KEITH  
; APPLICANT: STEWART, DAVID A.  
; APPLICANT: DANNER, DAVID B.  
; TITLE OF INVENTION: AN ANTIPROLIFERATIVE PROTEIN  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: 1100 NEW YORK AVENUE, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/612,674  
FILING DATE: 19901114  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SCOTT, WATSON T.  
REGISTRATION NUMBER: 26,581  
REFERENCE/DOCKET NUMBER: 5683/82332  
TELEPHONE: 202-861-3027  
TELEFAX: 202-822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-612-674-12

Query Match 100.0%; Score 19; DB 1; Length 93;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 18 VAEF 21

RESULT 29  
US-09-543-681A-5816

; Sequence 5816, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 5816  
; LENGTH: 98  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-5816

Query Match 100.0%; Score 19; DB 4; Length 98;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 21 VAEF 24

RESULT 30  
US-08-464-517-8  
; Sequence 8, Application US/08464517  
; Patent No. 5869640  
; GENERAL INFORMATION:  
; APPLICANT: BEACH, David H.  
; TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464,517  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/963,308  
FILING DATE: 16-OCT-1992  
APPLICATION NUMBER: US 07/989,178  
FILING DATE: 26-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/701,514  
FILING DATE: 16-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Matthew P. Vincent  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: MII-004C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-08-464-517-8

Query Match 100.0%; Score 19; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
b 13 VAEF 16

RESULT 31

S-08-246-361A-8  
Sequence 8, Application US/08246361A  
Patent No. 598582

## GENERAL INFORMATION:

APPLICANT: BEACH, David H.  
TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/246,361A  
FILING DATE: 19-MAY-1994  
CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/963,308  
FILING DATE: 16-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/888,178  
FILING DATE: 26-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/701,514  
FILING DATE: 16-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Matthew P. Vincent  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: MII-004C  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941

## INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

S-08-246-361A-8

Query Match 100.0%; Score 19; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
b 13 VAEF 16

RESULT 32

US-08-463-772-8  
Sequence 8, Application US/08463772  
Patent No. 6066501

## GENERAL INFORMATION:

APPLICANT: BEACH, David H.  
TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,772  
FILING DATE:  
CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/963,308  
FILING DATE: 16-OCT-1992  
APPLICATION NUMBER: US 07/888,178  
FILING DATE: 26-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/701,514  
FILING DATE: 16-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Matthew P. Vincent  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: MII-004C  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941

## INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-08-463-772-8

Query Match 100.0%; Score 19; DB 3; Length 100;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
b 13 VAEF 16

RESULT 33

PCT-US93-05000-8  
Sequence 8, Application PC/TUS9305000

## GENERAL INFORMATION:

APPLICANT: MITOTIX  
TITLE OF INVENTION: D-Type Cyclin and Uses Related Thereto  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02173

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/05000  
FILING DATE: 19930525  
CLASSIFICATION:  
PRIOR APPLICATION DATA: US/07/888,178  
FILING DATE: 26-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: CSHL91-02A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 618-861-9540  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
PCT-US93-05000-8

Query Match 100.0%; Score 19; DB 5; Length 100;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||  
Db 13 VAEF 16

RESULT 34  
US-08-580-988A-21  
Sequence 21, Application US/08580988A  
Patent No. 5856161  
GENERAL INFORMATION:  
APPLICANT: Agarwal et al.  
TITLE OF INVENTION: Tumor Necrosis Factor  
RECEPTOR-I-Associated Protein Kinase And Methods  
FOR ITS USE  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dr. Benjamin A. Adler  
STREET: 8011 Candle Lane  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 1.44 Mb floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word for Macintosh  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/580,988A  
FILING DATE: January 3, 1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benjamin Aaron Adler, Ph.D., J.D.  
REGISTRATION NUMBER: 35,423  
REFERENCE/DOCKET NUMBER: D5721CIP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-777-2321  
TELEFAX: 713-777-6908  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 101 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE:

DESCRIPTION: protein.  
HYPOTHETICAL: no  
ANTI-SENSE: no  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
US-08-580-988A-21

Query Match 100.0%; Score 19; DB 2; Length 101;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||  
Db 70 VAEF 73

RESULT 35  
US-09-540-236-3612  
Sequence 3612, Application US/09540236  
Patent No. 6673910  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATA  
FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.2005-001  
CURRENT APPLICATION NUMBER: US/09/540,236  
CURRENT FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 3840  
SEQ ID NO 3612  
LENGTH: 103  
TYPE: PRT  
ORGANISM: M.catarrhalis  
US-09-540-236-3612

Query Match 100.0%; Score 19; DB 4; Length 103;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||  
Db 65 VAEF 68

RESULT 36  
US-08-464-517-25  
Sequence 25, Application US/08464517  
Patent No. 5869640  
GENERAL INFORMATION:  
APPLICANT: BEACH, David H.  
TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464,517  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/963,308  
FILING DATE: 16-OCT-1992  
APPLICATION NUMBER: US 07/888,178  
FILING DATE: 26-MAY-1992  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/701,514  
FILING DATE: 16-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Matthew P. Vincent  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: MII-004C  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-464-517-25

Query Match 100.0%; Score 19; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 4e+02; 0; Indels 0;  
Matches 4; Conservative 0; Mismatches 0; Gaps 0;

QY 1 VAEF 4  
DB 66 VAEF 69

RESULT 37  
US-08-246-361A-25  
Sequence 25, Application US/08246361A  
Patent No. 5998582  
GENERAL INFORMATION:  
APPLICANT: BEACH, David H.  
TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THEREOF  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/246,361A  
FILING DATE: 19-MAY-1994  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/963,308  
FILING DATE: 16-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/888,178  
FILING DATE: 26-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/701,514  
FILING DATE: 16-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Matthew P. Vincent  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: MII-004C  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

; MOLECULE TYPE: peptide  
US-08-246-361A-25

Query Match 100.0%; Score 19; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 4e+02; 0; Indels 0;  
Matches 4; Conservative 0; Mismatches 0; Gaps 0;

QY 1 VAEF 4  
DB 66 VAEF 69

RESULT 38  
US-08-463-772-25  
Sequence 25, Application US/08463772  
Patent No. 6066501  
GENERAL INFORMATION:  
APPLICANT: BEACH, David H.  
TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THEREOF  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,772  
FILING DATE:  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/963,308  
FILING DATE: 16-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/888,178  
FILING DATE: 26-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/701,514  
FILING DATE: 16-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Matthew P. Vincent  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: MII-004C  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-463-772-25

Query Match 100.0%; Score 19; DB 3; Length 106;  
Best Local Similarity 100.0%; Pred. No. 4e+02; 0; Indels 0;  
Matches 4; Conservative 0; Mismatches 0; Gaps 0;

QY 1 VAEF 4  
DB 66 VAEF 69

RESULT 39  
US-09-489-039A-8721  
Sequence 8721, Application US/09489039A  
Patent No. 6610836

```

; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8721
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8721

Query Match      100.0%; Score 19; DB 4; Length 106;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 VAEF 4
Db      43 VAEF 46

RESULT 40
PCT-US93-05000-25
; Sequence 25, Application PC/TUS9305000
; GENERAL INFORMATION:
; APPLICANT: MITOTIX
; TITLE OF INVENTION: D-Type Cyclin and Uses Related Thereto
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05000
; FILING DATE: 19930525
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,178
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL91-02A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 616-861-9540
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: Protein
PCT-US93-05000-25

Query Match      100.0%; Score 19; DB 5; Length 106;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 VAEF 4
Db      66 VAEF 69
```

```

RESULT 41
US-09-489-039A-9915
; Sequence 9915, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9915
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9915

Query Match      100.0%; Score 19; DB 4; Length 115;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 VAEF 4
Db      50 VAEF 53

RESULT 42
US-09-252-991A-21614
; Sequence 21614, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21614
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21614

Query Match      100.0%; Score 19; DB 4; Length 122;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 VAEF 4
Db      85 VAEF 88

RESULT 43
US-09-489-039A-7200
; Sequence 7200, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
```

;; PRIOR APPLICATION NUMBER: US 60/117,747  
;; PRIOR FILING DATE: 1999-01-29  
;; NUMBER OF SEQ ID NOS: 14342  
;; SEQ ID NO 7200  
;; LENGTH: 123  
;; TYPE: PRT  
;; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-7200

Query Match  
Best Local Similarity 100.0%; Score 19; DB 4; Length 123;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 61 VAEF 64

RESULT 44  
US-08-529-055-53  
;; Sequence 53, Application US/08529055  
;; Patent No. 6592876  
;; GENERAL INFORMATION:  
;; APPLICANT: Biles, David E.  
;; APPLICANT: McDaniel, Larry S.  
;; APPLICANT: Swiatlo, Edwin  
;; APPLICANT: Yother, Janet  
;; APPLICANT: Brooks-Walter, Alexis  
;; TITLE OF INVENTION: Pneumococcal Genes, Portions  
;; TITLE OF INVENTION: Thereof, Expression Products  
;; TITLE OF INVENTION: Thereof, and Uses of Such Genes,  
;; TITLE OF INVENTION: Portions and Products  
;; NUMBER OF SEQUENCES: 73  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESS: Curtis, Morris & Safford, P.C.  
;; STREET: 530 Fifth Avenue  
;; CITY: New York  
;; STATE: NY  
;; COUNTRY: USA  
;; ZIP: 10036  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/529,055  
;; FILING DATE: 15-SEP-1995  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Frommer, William S.  
;; REGISTRATION NUMBER: 25,506  
;; REFERENCE/DOCKET NUMBER: 454312-2400  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 840-3333  
;; TELEFAX: (212) 840-0712  
;; INFORMATION FOR SEQ ID NO: 53:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 129 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
US-08-529-055-53

Query Match  
Best Local Similarity 100.0%; Score 19; DB 4; Length 129;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 7 VAEF 10

RESULT 45  
US-09-489-039A-12835  
;; Sequence 12835, Application US/09489039A  
;; Patent No. 6610836  
;; GENERAL INFORMATION:  
;; APPLICANT: Gary Bretton et. al  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
;; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 2709.2004001  
;; CURRENT APPLICATION NUMBER: US/09/489,039A  
;; CURRENT FILING DATE: 2000-01-27  
;; PRIOR APPLICATION NUMBER: US 60/117,747  
;; PRIOR FILING DATE: 1999-01-29  
;; NUMBER OF SEQ ID NOS: 14342  
;; SEQ ID NO 12835  
;; LENGTH: 143  
;; TYPE: PRT  
;; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-12835

Query Match  
Best Local Similarity 100.0%; Score 19; DB 4; Length 143;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 138 VAEF 141

RESULT 46  
US-09-198-452A-167  
;; Sequence 167, Application US/09198452A  
;; Patent No. 6559294  
;; GENERAL INFORMATION:  
;; APPLICANT: Griffiths, R.  
;; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment;  
;; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prev  
;; TITLE OF INVENTION: and treatment of infection  
;; FILE REFERENCE: 9710-003-999  
;; CURRENT APPLICATION NUMBER: US/09/198,452A  
;; CURRENT FILING DATE: 1998-11-24  
;; NUMBER OF SEQ ID NOS: 6849  
;; SEQ ID NO 167  
;; LENGTH: 145  
;; TYPE: PRT  
;; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-167

Query Match  
Best Local Similarity 100.0%; Score 19; DB 4; Length 145;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 31 VAEF 34

RESULT 47  
US-09-134-000C-4292  
;; Sequence 4292, Application US/09134000C  
;; Patent No. 6617156  
;; GENERAL INFORMATION:  
;; APPLICANT: Lynn Doucette-Stamm et al  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
;; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 032798-032  
;; CURRENT APPLICATION NUMBER: US/09/134,000C  
;; CURRENT FILING DATE: 1998-08-13  
;; PRIOR APPLICATION NUMBER: US 60/055,778  
;; PRIOR FILING DATE: 1997-08-15  
;; NUMBER OF SEQ ID NOS: 6812  
;; SOFTWARE: PatentIn version 3.1

/ SEQ ID NO 4292  
/ LENGTH: 145  
/ TYPE: PRT  
/ ORGANISM: Enterococcus faecalis  
US-09-134-000C-4292

Query Match 100.0%; Score 19; DB 4; Length 145;  
Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 113 VAEF 116

RESULT 48  
US-08-460-694-3  
Sequence 3, Application US/08460694  
Patent No. 5858655

/ GENERAL INFORMATION:  
/ APPLICANT: Arnold, Andrew  
/ TITLE OF INVENTION: Pradi Cyclin and its cDNA  
/ NUMBER OF SEQUENCES: 8  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
/ STREET: 1100 New York Avenue, N.W., Suite 600  
/ CITY: Washington  
/ STATE: DC  
/ COUNTRY: USA  
/ ZIP: 20005

/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/460,694  
/ FILING DATE: 02-JUN-1995

/ CLASSIFICATION: 435  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: McConathy, Evelyn H.

/ REGISTRATION NUMBER: 35,279  
/ REFERENCE/DOCKET NUMBER: 0609.4070002  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 202-371-2600  
/ TELEFAX: 202-371-2540

/ INFORMATION FOR SEQ ID NO: 3:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 150 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: not relevant  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: peptide  
US-08-460-694-3

Query Match 100.0%; Score 19; DB 2; Length 150;  
Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 66 VAEF 69

RESULT 49  
US-08-460-744-3  
Sequence 3, Application US/08460744  
Patent No. 6107541

/ GENERAL INFORMATION:  
/ APPLICANT: Arnold, Andrew  
/ TITLE OF INVENTION: Pradi Cyclin and its cDNA  
/ NUMBER OF SEQUENCES: 8  
/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
/ STREET: 1100 New York Avenue, N.W., Suite 600  
/ CITY: Washington  
/ STATE: DC  
/ COUNTRY: USA  
/ ZIP: 20005

/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/460,744  
/ FILING DATE: 02-JUN-1995

/ CLASSIFICATION: 435  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: McConathy, Evelyn H.  
/ REGISTRATION NUMBER: 35,279  
/ REFERENCE/DOCKET NUMBER: 0609.4070005  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 202-371-2600  
/ TELEFAX: 202-371-2540

/ INFORMATION FOR SEQ ID NO: 3:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 150 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: not relevant  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: peptide  
US-08-460-744-3

Query Match 100.0%; Score 19; DB 3; Length 150;  
Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 66 VAEF 69

RESULT 50

US-07-667-711B-3  
Sequence 3, Application US/07667711B  
Patent No. 6110700

/ GENERAL INFORMATION:  
/ APPLICANT: ARNOLD, ANDREW  
/ TITLE OF INVENTION: Pradi Cyclin and its cDNA  
/ NUMBER OF SEQUENCES: 8  
/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
/ STREET: 1100 NEW YORK AVE., NW, SUITE 600  
/ CITY: WASHINGTON  
/ STATE: DC  
/ COUNTRY: USA  
/ ZIP: 20005

/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/07/667,711B  
/ FILING DATE: 11-MAR-1991

/ CLASSIFICATION: 435  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: MCPHAIL, DONALD R.  
/ REGISTRATION NUMBER: 35,811  
/ REFERENCE/DOCKET NUMBER: 0609.4070000  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (202) 371-2600  
/ TELEFAX: (202) 371-2540  
/ INFORMATION FOR SEQ ID NO: 3:  
/ SEQUENCE CHARACTERISTICS:

LENGTH: 150 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
-07-667-711B-3

Query Match 100.0%; Score 19; DB 3; Length 150;  
Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
|||||  
66 VAEF 69

SULT 51  
-08-679-493A-208  
Sequence 208, Application US/08679493A  
Patent No. 6303295  
GENERAL INFORMATION:  
APPLICANT: Taylor, Ethan W.  
TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS  
FILE REFERENCE: 55-95  
CURRENT APPLICATION NUMBER: US/08/679,493A  
CURRENT FILING DATE: 1996-07-12  
PRIOR APPLICATION NUMBER: 60/001203  
PRIOR FILING DATE: 1995-07-14  
PRIOR APPLICATION NUMBER: 60/003,112  
PRIOR FILING DATE: 1995-09-01  
NUMBER OF SEQ ID NOS: 216  
SOFTWARE: Patent in ver. 2.0  
SEQ ID NO 208  
LENGTH: 152  
TYPE: PRT  
ORGANISM: blueshark  
-08-679-493A-208

Query Match 100.0%; Score 19; DB 4; Length 152;  
Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
|||||  
92 VAEF 95

SULT 52  
-09-134-000C-4087  
Sequence 4087, Application US/09134000C  
Patent No. 6617156  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 032796-032  
CURRENT APPLICATION NUMBER: US/09/134,000C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/055,778  
PRIOR FILING DATE: 1997-08-15  
NUMBER OF SEQ ID NOS: 6812  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 4087  
LENGTH: 153  
TYPE: PRT  
ORGANISM: Enterococcus faecalis  
-09-134-000C-4087

Query Match 100.0%; Score 19; DB 4; Length 153;  
Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4

Db 101 VAEF 104  
|||||

RESULT 53  
US-09-252-991A-20805  
Sequence 20805, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 20805  
LENGTH: 156  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20805

Query Match 100.0%; Score 19; DB 4; Length 156;  
Best Local Similarity 100.0%; Pred. No. 6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||||  
Db 84 VAEF 87

RESULT 54  
US-09-370-838-203  
Sequence 203, Application US/09370838  
Patent No. 6444425  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Mohamath, Roadoh  
APPLICANT: Secrist, Heather  
TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF  
TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE  
FILE REFERENCE: 210121.475C1  
CURRENT APPLICATION NUMBER: US/09/370,838  
CURRENT FILING DATE: 1999-08-09  
EARLIER APPLICATION NUMBER: US 09/285,323  
EARLIER FILING DATE: 1999-04-02  
NUMBER OF SEQ ID NOS: 289  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 203  
LENGTH: 164  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-370-838-203

Query Match 100.0%; Score 19; DB 4; Length 164;  
Best Local Similarity 100.0%; Pred. No. 6.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||||  
Db 100 VAEF 103

RESULT 55  
US-09-540-236-2222  
Sequence 2222, Application US/09540236  
Patent No. 6673910  
GENERAL INFORMATION:

; APPLICANT: Gary L. Bretton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS  
; FILE REFERENCE: 2709.2005-001  
; CURRENT APPLICATION NUMBER: US/09/540,236  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 3840  
; SEQ ID NO 2222  
; LENGTH: 166  
; TYPE: PRT  
; ORGANISM: M.catarrahals  
US-09-540-236-2222

Query Match 100.0%; Score 19; DB 4; Length 166;  
Best Local Similarity 100.0%; Pred. No. 6.4e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 97 VAEF 100

RESULT 56  
US-08-471-058-16  
; Sequence 16, Application US/08471058  
; Patent No. 5770443  
; GENERAL INFORMATION:  
; APPLICANT: Kiefer, Michael C.  
; APPLICANT: Barr, Philip J.  
; TITLE OF INVENTION: NOVEL APOPTOSIS MODULATING  
; TITLE OF INVENTION: PROTEINS, DNA ENCODING THE PROTEINS AND METHODS OF USE  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,058  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
Prior Application Data:  
APPLICATION NUMBER: 08/320,157  
FILING DATE: 07-OCT-1994  
APPLICATION NUMBER: 08/160,067  
FILING DATE: 30-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lehnhardt, Susan K  
REGISTRATION NUMBER: 33,943  
REFERENCE/DOCKET NUMBER: 23647-20007.12  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-813-5600  
TELEFAX: 415-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 172 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-471-058-16

Query Match 100.0%; Score 19; DB 1; Length 172;  
Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 122 VAEF 125

RESULT 57  
US-08-690-095-3  
; Sequence 3, Application US/08690095  
; Patent No. 5792648  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL HUMAN MACROPHAGE ANTIGEN  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/690,095  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PP-0110 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 172 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 293274  
US-08-690-095-3

Query Match 100.0%; Score 19; DB 1; Length 172;  
Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
DB 122 VAEF 125

RESULT 58  
US-08-471-057-16  
; Sequence 16, Application US/08471057  
; Patent No. 6015687  
; GENERAL INFORMATION:  
; APPLICANT: KIEFER, MICHAEL C.  
; APPLICANT: BARR, PHILIP J.  
; TITLE OF INVENTION: NOVEL APOPTOSIS-MODULATING PROTEINS, DNA  
; TITLE OF INVENTION: ENCODING THE PROTEINS AND METHODS OF USE THEREOF  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto

STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/471,057  
APPLICATION NUMBER: US/08/471,057  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/320,157  
FILING DATE: 07-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: LEHNHARDT, SUSAN K.  
REGISTRATION NUMBER: 33,943  
REFERENCE/DOCKET NUMBER: 23647-20007.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 172 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-471-057-16

Query Match 100.0%; Score 19; DB 3; Length 172;  
Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
122 VAEF 125

RESULT 59  
US-09-113-789-3  
Sequence 3, Application US/09113789  
Patent No. 6034219  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Au-Young, Janice  
APPLICANT: Goli, Sulva K.  
TITLE OF INVENTION: NOVEL HUMAN MACROPHAGE ANTIGEN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/113,789  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/690,095  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0110 US  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 172 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 293274  
US-09-113-789-3

Query Match 100.0%; Score 19; DB 3; Length 172;  
Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
122 VAEF 125

RESULT 60  
US-08-470-865-16  
Sequence 16, Application US/08470865  
Patent No. 6586395  
GENERAL INFORMATION:  
APPLICANT: KIEFER, MICHAEL C.  
APPLICANT: BARR, PHILIP J.  
TITLE OF INVENTION: NOVEL APOPTOSIS-MODULATING PROTEINS, DNA  
TITLE OF INVENTION: ENCODING THE PROTEINS AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,865  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/320,157  
FILING DATE: 07-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: LEHNHARDT, SUSAN K.  
REGISTRATION NUMBER: 33,943  
REFERENCE/DOCKET NUMBER: 23647-20007.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 172 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-470-865-16

Query Match 100.0%; Score 19; DB 4; Length 172;  
Best Local Similarity 100.0%; Pred. No. 6.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4

```

; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9869

Query Match      100.0%; Score 19; DB 4; Length 174;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 77 VAEF 80

RESULT 61
US-08-193-977-4
; Sequence 4, Application US/08193977
; Patent No. 5625031
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, KEVIN R.
; APPLICANT: COLEMAN, KEVIN G.
; TITLE OF INVENTION: PEPTIDE INHIBITORS OF THE P33CDK2 AND
; TITLE OF INVENTION: P34CDK2 CELL CYCLE REGULATORY KINASES AND HUMAN
; TITLE OF INVENTION: PAPILLOMAVIRUS E7 ONCOPROTEIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESS: REED & ROBINS
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/193,977
; FILING DATE: 08-FEB-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0016
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 173 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-193-977-4

Query Match      100.0%; Score 19; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 6.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 100 VAEF 103

RESULT 62
US-09-489-039A-9869
; Sequence 9869, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9869
; LENGTH: 174
; TYPE: PRT

; ORGANISM: Acinetobacter baumannii
US-09-328-352-6739

Query Match      100.0%; Score 19; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 8 VAEF 11

RESULT 63
US-08-737-980-2
; Sequence 2, Application US/08737980
; Patent No. 5843773
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Apoptosis Regulating Gene
; NUMBER OF SEQUENCES: 2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737,980
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: KR 1995-6266
; FILING DATE: 24-MAR-1995
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 175 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-737-980-2

Query Match      100.0%; Score 19; DB 2; Length 175;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 122 VAEF 125

RESULT 64
US-09-328-352-6739
; Sequence 6739, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6739
; LENGTH: 180
; TYPE: PRT

; ORGANISM: Acinetobacter baumannii
US-09-328-352-6739

Query Match      100.0%; Score 19; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4
DB 8 VAEF 11

```

SULT 65  
-09-252-991A-24341  
Sequence 24341, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 24341  
LENGTH: 185  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
-09-252-991A-24341  
Query Match 100.0%; Score 19; DB 4; Length 185;  
Best Local Similarity 100.0%; Pred. No. 7.2e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 VAEF 4  
4 VAEF 7  
SULT 66  
-09-543-681A-7313  
Sequence 7313, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 7313  
LENGTH: 189  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
-09-543-681A-7313  
Query Match 100.0%; Score 19; DB 4; Length 189;  
Best Local Similarity 100.0%; Pred. No. 7.4e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 VAEF 4  
73 VAEF 76  
SULT 67  
-08-816-241-1  
Sequence 1, Application US/08816241  
Patent No. 5804185  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,241  
FILING DATE: Filed Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTUT09  
CLONE: 1646823  
US-08-816-241-1  
Query Match 100.0%; Score 19; DB 1; Length 190;  
Best Local Similarity 100.0%; Pred. No. 7.4e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
DB 104 VAEF 107  
RESULT 68  
US-09-128-395-1  
Sequence 1, Application US/09128395  
Patent No. 6087108  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,395  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/816,241  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
 REGISTRATION NUMBER: 36,749  
 REFERENCE/DOCKET NUMBER: PF-0239 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-845-0555  
 TELEFAX: 415-845-4166  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 190 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: PROSTUT09  
 CLONE: 1646823  
 S-09-128-395-1

Query Match 100.0%; Score 19; DB 3; Length 190;  
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 b 104 VAEF 107

RESULT 69  
 US-09-134-000C-3800  
 Sequence 3800, Application US/09134000C  
 Patent No. 6617156  
 GENERAL INFORMATION:  
 APPLICANT: Lynn Doucette-Stamm et al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 FILE REFERENCE: 032796-032  
 CURRENT APPLICATION NUMBER: US/09/134,000C  
 CURRENT FILING DATE: 1998-08-13  
 PRIOR APPLICATION NUMBER: US 60/055,778  
 PRIOR FILING DATE: 1997-08-15  
 NUMBER OF SEQ ID NOS: 6812  
 SOFTWARE: Patent in version 3.1  
 SEQ ID NO 3800  
 LENGTH: 194  
 TYPE: PRT  
 ORGANISM: Enterococcus faecalis  
 US-09-134-000C-3800

Query Match 100.0%; Score 19; DB 4; Length 194;  
 Best Local Similarity 100.0%; Pred. No. 7.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 b 115 VAEF 118

RESULT 70  
 US-09-232-200-85  
 Sequence 85, Application US/09232200A  
 Patent No. 6289313  
 GENERAL INFORMATION:  
 APPLICANT: Stahl, Andreas  
 APPLICANT: Hirsch, David J.  
 APPLICANT: Lodish, Harvey F.  
 APPLICANT: Gimeno, Ruth E.  
 APPLICANT: Tartaglia, Louis A.  
 TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
 FILE REFERENCE: WH197-21p3MB  
 CURRENT APPLICATION NUMBER: US/09/232,200A  
 CURRENT FILING DATE: 1999-01-14  
 EARLIER APPLICATION NUMBER: 60/071,374  
 EARLIER FILING DATE: 1998-01-15  
 EARLIER APPLICATION NUMBER: 60/093,491

NAME: Billings, Lucy J.  
 REGISTRATION NUMBER: 36,749  
 REFERENCE/DOCKET NUMBER: PF-0239 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-845-0555  
 TELEFAX: 415-845-4166  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 199  
 TYPE: PRT  
 ORGANISM: Aspergillus nidulans  
 US-09-232-200-85

Query Match 100.0%; Score 19; DB 3; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 b 99 VAEF 102

RESULT 71  
 US-09-232-197-85  
 Sequence 85, Application US/09232197A  
 Patent No. 6300096  
 GENERAL INFORMATION:  
 APPLICANT: Stahl, Andreas  
 APPLICANT: Hirsch, David J.  
 APPLICANT: Lodish, Harvey F.  
 APPLICANT: Gimeno, Ruth E.  
 APPLICANT: Tartaglia, Louis A.  
 TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
 FILE REFERENCE: WH197-21p3MA  
 CURRENT APPLICATION NUMBER: US/09/232,197A  
 CURRENT FILING DATE: 1999-01-14  
 EARLIER APPLICATION NUMBER: 60/071,374  
 EARLIER FILING DATE: 1998-01-15  
 EARLIER APPLICATION NUMBER: 60/093,491  
 EARLIER FILING DATE: 1998-07-20  
 EARLIER APPLICATION NUMBER: 60/110,941  
 EARLIER FILING DATE: 1998-12-04  
 NUMBER OF SEQ ID NOS: 105  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 85  
 LENGTH: 199  
 TYPE: PRT  
 ORGANISM: Aspergillus nidulans  
 US-09-232-197-85

Query Match 100.0%; Score 19; DB 4; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 b 99 VAEF 102

RESULT 72  
 US-09-232-201-85  
 Sequence 85, Application US/09232201A  
 Patent No. 6348321  
 GENERAL INFORMATION:  
 APPLICANT: Stahl, Andreas  
 APPLICANT: Hirsch, David J.  
 APPLICANT: Lodish, Harvey F.  
 APPLICANT: Gimeno, Ruth E.  
 APPLICANT: Tartaglia, Louis A.  
 TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
 FILE REFERENCE: WH197-21p3MC  
 CURRENT APPLICATION NUMBER: US/09/232,201A  
 CURRENT FILING DATE: 1999-01-14  
 EARLIER APPLICATION NUMBER: 60/071,374  
 EARLIER FILING DATE: 1998-01-15  
 EARLIER APPLICATION NUMBER: 60/093,491

EARLIER APPLICATION NUMBER: 60/093,491  
EARLIER FILING DATE: 1998-07-20  
EARLIER APPLICATION NUMBER: 60/110,941  
EARLIER FILING DATE: 1998-12-04  
NUMBER OF SEQ ID NOS: 105  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 85  
LENGTH: 199  
TYPE: PRT  
ORGANISM: Aspergillus nidulans  
;-09-232-201-85

Query Match 100.0%; Score 19; DB 4; Length 199;  
Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 VAEF 4  
99 VAEF 102

RESULT 73  
;-09-232-195-85  
Sequence 85, Application US/09232195A  
Patent No. 6657049  
GENERAL INFORMATION:  
APPLICANT: Stahl, Andreas  
APPLICANT: Hirsch, David J.  
APPLICANT: Lodish, Harvey F.  
APPLICANT: Gimeno, Ruth E.  
APPLICANT: Tartaglia, Louis A.  
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS  
FILE REFERENCE: WH197-21p3MD  
CURRENT APPLICATION NUMBER: US/09/232,195A  
CURRENT FILING DATE: 1999-01-04  
EARLIER APPLICATION NUMBER: 60/071,374  
EARLIER FILING DATE: 1998-01-15  
EARLIER APPLICATION NUMBER: 60/093,491  
EARLIER FILING DATE: 1998-07-20  
EARLIER APPLICATION NUMBER: 60/110,941  
EARLIER FILING DATE: 1998-12-04  
NUMBER OF SEQ ID NOS: 105  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 85  
LENGTH: 199  
TYPE: PRT  
ORGANISM: Aspergillus nidulans  
;-09-232-195-85

Query Match 100.0%; Score 19; DB 4; Length 199;  
Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 VAEF 4  
99 VAEF 102

RESULT 74  
;-09-107-532A-4612  
Sequence 4612, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESS: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA

ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GFC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4612:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 203 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...203  
SEQUENCE DESCRIPTION: SEQ ID NO: 4612:  
US-09-107-532A-4612  
Query Match 100.0%; Score 19; DB 4; Length 203;  
Best Local Similarity 100.0%; Pred. No. 7.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
DB 63 VAEF 66  
RESULT 75  
US-08-684-024-1  
Sequence 1, Application US/08684024  
Patent No. 5834298  
GENERAL INFORMATION:  
APPLICANT: Benzra, Robert  
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/684,024  
FILING DATE: 19-JUL-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0526

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 205 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

IS-08-684-024-1

Query Match 100.0%; Score 19; DB 2; Length 205;

Best Local Similarity 100.0%; Pred. No. 8e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4

20 VAEF 23

earch completed: May 24, 2004, 17:43:22

Ob time : 14 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: May 24, 2004, 17:34:23 ; Search time 31.7143 Seconds  
(without alignments)

35.179 Million cell updates/sec

tie: " US-09-594-978A-2

fect score: 19

quence: 1 VAEF 4

oring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

arched: 1149313 seqs, 278921704 residues

tal number of hits satisfying chosen parameters: 1149313

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

tabase : Published Applications AA: \*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

sult	No.	Score	Query	Match	Length	DB	ID	Description
1	19	100.0	8	9	US-09-791-378-343			Sequence 343, App
2	19	100.0	8	10	US-09-791-393-206			Sequence 206, App
3	19	100.0	8	10	US-09-791-389-206			Sequence 206, App
4	19	100.0	9	14	US-10-128-711-86			Sequence 86, Appl
5	19	100.0	9	14	US-10-128-711-145			Sequence 145, App
6	19	100.0	10	12	US-09-508-943A-137			Sequence 197, App
7	19	100.0	12	12	US-09-508-943A-136			Sequence 196, App
8	19	100.0	12	14	US-10-032-818-34			Sequence 34, Appl
9	19	100.0	13	14	US-10-239-313A-379			Sequence 379, App
10	19	100.0	15	14	US-10-119-528-7			Sequence 7, Appl
11	19	100.0	20	10	US-09-171-432A-62			Sequence 62, Appl
12	19	100.0	20	10	US-09-171-432A-63			Sequence 63, Appl
13	19	100.0	25	10	US-09-171-432A-65			Sequence 65, Appl
14	19	100.0	28	9	US-09-864-761-47201			Sequence 47201, A
15	19	100.0	28	12	US-10-424-599-216297			Sequence 216297,

16	19	100.0	35	14	US-10-043-344-59			Sequence 69, Appl
17	19	100.0	37	15	US-10-242-355-570			Sequence 570, App
18	19	100.0	37	15	US-10-430-752A-49			Sequence 49, Appl
19	19	100.0	38	15	US-10-430-752A-45			Sequence 45, Appl
20	19	100.0	41	15	US-10-372-003A-71			Sequence 71, Appl
21	19	100.0	43	9	US-09-864-761-37202			Sequence 37202, A
22	19	100.0	43	12	US-10-424-599-183032			Sequence 183032,
23	19	100.0	45	12	US-10-424-599-171318			Sequence 171318,
24	19	100.0	48	9	US-09-864-761-38166			Sequence 38166, A
25	19	100.0	52	12	US-10-424-599-205632			Sequence 205632,
26	19	100.0	55	12	US-10-424-599-192247			Sequence 192247,
27	19	100.0	55	12	US-10-424-599-233502			Sequence 233502,
28	19	100.0	55	15	US-10-372-003A-63			Sequence 63, Appl
29	19	100.0	55	15	US-10-372-003A-67			Sequence 67, Appl
30	19	100.0	56	14	US-10-213-329-4			Sequence 4, Appl
31	19	100.0	56	14	US-10-153-185-4			Sequence 4, Appl
32	19	100.0	56	14	US-10-219-561-4			Sequence 4, Appl
33	19	100.0	57	12	US-10-424-599-215734			Sequence 215734,
34	19	100.0	57	12	US-10-424-599-260762			Sequence 260762,
35	19	100.0	57	15	US-10-372-003A-75			Sequence 75, Appl
36	19	100.0	58	12	US-10-424-599-214930			Sequence 214930,
37	19	100.0	60	9	US-09-864-761-45291			Sequence 45291, A
38	19	100.0	68	9	US-09-393-634-80			Sequence 80, Appl
39	19	100.0	68	11	US-09-864-408A-40			Sequence 40, Appl
40	19	100.0	68	12	US-10-364-861-80			Sequence 80, Appl
41	19	100.0	68	14	US-10-383-982-80			Sequence 80, Appl
42	19	100.0	69	9	US-09-864-761-45716			Sequence 45716, A
43	19	100.0	69	9	US-09-738-626-5596			Sequence 5596, Ap
44	19	100.0	69	12	US-10-424-599-185385			Sequence 185385,
45	19	100.0	70	14	US-10-023-386-28645			Sequence 28645, A
46	19	100.0	71	12	US-10-424-599-202605			Sequence 202605,
47	19	100.0	71	12	US-10-424-599-274981			Sequence 274981,
48	19	100.0	71	14	US-10-029-386-30348			Sequence 30348, A
49	19	100.0	72	14	US-10-029-386-29320			Sequence 29320, A
50	19	100.0	74	14	US-10-214-188-10			Sequence 10, Appl
51	19	100.0	74	14	US-10-156-761-9071			Sequence 9071, Ap
52	19	100.0	75	12	US-10-424-599-156128			Sequence 156128,
53	19	100.0	75	12	US-10-424-599-261424			Sequence 261424,
54	19	100.0	76	12	US-10-424-599-253204			Sequence 253204,
55	19	100.0	77	12	US-10-424-599-211908			Sequence 211908,
56	19	100.0	77	12	US-10-335-977-7493			Sequence 7493, Ap
57	19	100.0	77	14	US-10-029-386-32709			Sequence 32709, A
58	19	100.0	78	12	US-10-424-599-162039			Sequence 162039,
59	19	100.0	78	12	US-10-424-599-262701			Sequence 262701,
60	19	100.0	79	12	US-10-363-616-368			Sequence 368, App
61	19	100.0	80	12	US-10-424-599-216037			Sequence 216037,
62	19	100.0	82	15	US-10-367-980A-14			Sequence 14, Appl
63	19	100.0	84	11	US-09-864-408A-5282			Sequence 5282, Ap
64	19	100.0	84	14	US-10-078-080-125			Sequence 125, App
65	19	100.0	85	12	US-10-424-599-155636			Sequence 155636,
66	19	100.0	85	12	US-10-424-599-201853			Sequence 201853,
67	19	100.0	89	9	US-09-815-242-11224			Sequence 11224, A
68	19	100.0	89	9	US-09-815-242-11245			Sequence 11245, A
69	19	100.0	89	12	US-10-283-122A-58489			Sequence 58489, A
70	19	100.0	89	12	US-10-283-122A-58530			Sequence 58530, A
71	19	100.0	89	12	US-10-283-122A-66960			Sequence 66960, A
72	19	100.0	89	12	US-10-282-122A-68774			Sequence 68774, A
73	19	100.0	89	12	US-10-424-599-180307			Sequence 180307,
74	19	100.0	89	14	US-10-097-111-360			Sequence 360, App
75	19	100.0	90	12	US-10-424-599-245830			Sequence 245830,
76	19	100.0	93	12	US-10-424-599-200533			Sequence 200533,
77	19	100.0	93	12	US-10-424-599-204269			Sequence 204269,
78	19	100.0	94	15	US-10-372-003A-79			Sequence 79, Appl
79	19	100.0	99	15	US-10-264-043-2775			Sequence 2775, Ap
80	19	100.0	100	12	US-10-335-977-7495			Sequence 7495, Ap
81	19	100.0	101	12	US-10-335-977-7494			Sequence 7494, Ap
82	19	100.0	101	12	US-10-424-599-196121			Sequence 196121,
83	19	100.0	104	12	US-10-424-599-245993			Sequence 245993,
84	19	100.0	104	16	US-10-294-445-30			Sequence 30, Appl
85	19	100.0	105	14	US-10-291-851-90			Sequence 90, Appl
86	19	100.0	107	9	US-09-738-626-4520			Sequence 4520, Ap
87	19	100.0	109	12	US-10-363-616-367			Sequence 367, App
88	19	100.0	109	15	US-10-264-237-1538			Sequence 1538, Ap

89	19	100.0	110	12	US-10-424-599-144932	Sequence 144932, App	162	19	100.0	189	12	US-10-424-599-208651	Sequence 208651, App
90	19	100.0	110	12	US-10-424-599-257728	Sequence 257728, App	163	19	100.0	189	14	US-10-238-075-1037	Sequence 1037, App
91	19	100.0	111	12	US-10-424-599-1927406	Sequence 1927406, App	164	19	100.0	191	12	US-10-282-122A-77681	Sequence 77681, App
92	19	100.0	112	12	US-10-424-599-189066	Sequence 189066, App	165	19	100.0	192	12	US-10-424-599-269537	Sequence 269537, App
93	19	100.0	115	12	US-10-424-599-222263	Sequence 222263, App	166	19	100.0	193	9	US-09-888-911-4	Sequence 4, Appli
94	19	100.0	117	12	US-10-424-599-206883	Sequence 206883, App	167	19	100.0	193	12	US-10-425-114-43750	Sequence 43750, A
95	19	100.0	117	12	US-10-424-599-221584	Sequence 221584, App	168	19	100.0	193	14	US-10-156-761-13355	Sequence 13355, A
96	19	100.0	118	12	US-10-424-599-271082	Sequence 271082, App	169	19	100.0	193	14	US-10-377-072-38	Sequence 38, Appl
97	19	100.0	119	12	US-10-424-599-240644	Sequence 240644, App	170	19	100.0	195	12	US-10-424-599-231551	Sequence 231551, App
98	19	100.0	119	14	US-10-263-568-12	Sequence 12, Appl	171	19	100.0	195	12	US-10-424-599-238939	Sequence 238939, App
99	19	100.0	120	12	US-10-424-599-189718	Sequence 189718, App	172	19	100.0	197	15	US-10-369-493-3234	Sequence 3234, App
100	19	100.0	120	12	US-10-424-599-276785	Sequence 276785, App	173	19	100.0	197	15	US-09-966-880A-2	Sequence 2, Appli
101	19	100.0	122	11	US-09-864-408A-526	Sequence 526, App	174	19	100.0	198	9	US-10-425-114-47425	Sequence 47425, A
102	19	100.0	122	12	US-10-424-599-173522	Sequence 173522, App	175	19	100.0	199	12	US-10-405-877-85	Sequence 85, Appl
103	19	100.0	127	12	US-10-424-599-178767	Sequence 178767, App	176	19	100.0	199	12	US-10-424-599-176742	Sequence 176742, App
104	19	100.0	129	16	US-10-299-636-68	Sequence 68, Appl	177	19	100.0	200	12	US-10-424-599-220990	Sequence 220990, App
105	19	100.0	130	12	US-10-424-599-184879	Sequence 184879, App	178	19	100.0	200	12	US-10-276-162-2	Sequence 2, Appli
106	19	100.0	130	12	US-10-424-599-199911	Sequence 199911, App	179	19	100.0	201	15	US-10-085-198-134	Sequence 134, App
107	19	100.0	131	9	US-09-738-626-5204	Sequence 5204, App	180	19	100.0	201	15	US-10-282-122A-60863	Sequence 60863, A
108	19	100.0	132	12	US-10-243-552-332	Sequence 332, App	181	19	100.0	202	12	US-10-424-599-231852	Sequence 231852, App
109	19	100.0	133	9	US-09-864-761-33708	Sequence 33708, App	182	19	100.0	204	12	US-10-424-599-265941	Sequence 265941, App
110	19	100.0	133	12	US-10-424-599-158331	Sequence 158331, App	183	19	100.0	204	12	US-10-634-548-29	Sequence 29, Appl
111	19	100.0	134	9	US-09-815-242-10513	Sequence 10513, App	184	19	100.0	204	12	US-10-634-548-29	Sequence 31, Appl
112	19	100.0	135	12	US-10-424-599-271153	Sequence 271153, App	185	19	100.0	204	12	US-09-932-217-1108	Sequence 1108, App
113	19	100.0	135	12	US-10-425-114-53378	Sequence 53378, App	186	19	100.0	205	9	US-10-296-115-1330	Sequence 1330, App
114	19	100.0	138	12	US-10-112-944-312	Sequence 912, App	187	19	100.0	205	12	US-10-335-977-7988	Sequence 7988, App
115	19	100.0	139	14	US-10-156-761-8842	Sequence 8842, App	188	19	100.0	205	12	US-10-335-977-7988	Sequence 1108, App
116	19	100.0	139	15	US-10-385-415-120	Sequence 120, App	189	19	100.0	205	13	US-10-025-380-1108	Sequence 44800, A
117	19	100.0	140	15	US-10-108-260A-3390	Sequence 3390, App	190	19	100.0	208	12	US-10-282-122A-44800	Sequence 13562, A
118	19	100.0	145	15	US-10-385-415-119	Sequence 119, App	191	19	100.0	208	14	US-10-156-761-13562	Sequence 72561, A
119	19	100.0	145	13	US-10-013-379-11	Sequence 11, Appl	192	19	100.0	209	12	US-10-282-122A-72561	Sequence 267539, App
120	19	100.0	145	15	US-10-289-762-167	Sequence 167, App	193	19	100.0	209	12	US-10-424-599-267539	Sequence 130, App
121	19	100.0	146	12	US-10-289-132A-70260	Sequence 70260, App	194	19	100.0	212	14	US-10-127-032-130	Sequence 17, Appl
122	19	100.0	146	15	US-10-138-784-87	Sequence 87, Appl	195	19	100.0	213	9	US-09-943-671-17	Sequence 3673, App
123	19	100.0	148	12	US-10-424-599-143742	Sequence 143742, App	196	19	100.0	213	9	US-09-738-626-3673	Sequence 3673, App
124	19	100.0	148	12	US-10-424-599-219032	Sequence 219032, App	197	19	100.0	213	14	US-10-043-487-364	Sequence 164, App
125	19	100.0	152	12	US-10-424-599-283972	Sequence 283972, App	198	19	100.0	213	14	US-10-405-877-17	Sequence 17, Appl
126	19	100.0	152	13	US-10-016-157A-151	Sequence 151, App	199	19	100.0	215	14	US-10-106-698-5887	Sequence 5887, App
127	19	100.0	156	12	US-10-016-157A-151	Sequence 254161, App	200	19	100.0	215	9	US-09-738-626-5708	Sequence 5708, App
128	19	100.0	157	14	US-10-238-075-357	Sequence 357, App	201	19	100.0	215	12	US-10-326-671-69	Sequence 62976, A
129	19	100.0	159	12	US-10-424-599-256669	Sequence 256669, App	202	19	100.0	215	14	US-10-282-122A-62976	Sequence 69, Appl
130	19	100.0	163	12	US-10-424-599-169320	Sequence 169320, App	203	19	100.0	217	12	US-10-343-552-890	Sequence 890, App
131	19	100.0	164	9	US-09-738-973-203	Sequence 203, App	204	19	100.0	217	12	US-10-282-122A-68326	Sequence 88326, App
132	19	100.0	164	9	US-09-854-133-203	Sequence 203, App	205	19	100.0	218	12	US-10-282-122A-78537	Sequence 78537, A
133	19	100.0	164	11	US-09-864-408A-1148	Sequence 1148, App	206	19	100.0	218	12	US-10-424-599-154224	Sequence 154224, App
134	19	100.0	164	14	US-10-144-649A-203	Sequence 203, App	207	19	100.0	219	14	US-10-156-761-9995	Sequence 9995, App
135	19	100.0	165	12	US-10-424-599-148999	Sequence 148999, App	208	19	100.0	219	14	US-10-424-599-162041	Sequence 162041, App
136	19	100.0	165	12	US-10-424-599-234773	Sequence 234773, App	209	19	100.0	220	12	US-10-384-060-76	Sequence 76, Appl
137	19	100.0	167	9	US-09-925-300-981	Sequence 981, App	210	19	100.0	220	12	US-10-424-599-268598	Sequence 268598, App
138	19	100.0	168	12	US-10-424-599-198953	Sequence 198953, App	211	19	100.0	221	12	US-10-424-599-1639	Sequence 1639, App
139	19	100.0	172	12	US-10-425-114-57575	Sequence 57575, A	212	19	100.0	221	14	US-10-424-599-157122	Sequence 157122, App
140	19	100.0	172	14	US-10-101-482-16	Sequence 16, Appl	213	19	100.0	222	12	US-10-425-114-36974	Sequence 36974, A
141	19	100.0	173	12	US-10-424-599-158029	Sequence 158029, App	214	19	100.0	222	12	US-10-425-114-52549	Sequence 52549, A
142	19	100.0	175	12	US-10-335-977-5868	Sequence 5868, App	215	19	100.0	222	12	US-10-425-114-52549	Sequence 4621, App
143	19	100.0	175	14	US-10-233-984-12	Sequence 12, Appl	216	19	100.0	224	9	US-09-738-626-4621	Sequence 151189, App
144	19	100.0	175	14	US-10-156-761-12575	Sequence 12575, A	217	19	100.0	224	12	US-10-424-599-151189	Sequence 151189, App
145	19	100.0	175	14	US-10-247-671-184	Sequence 184, App	218	19	100.0	224	12	US-10-394-749-1944	Sequence 1944, App
146	19	100.0	176	9	US-09-864-761-33919	Sequence 33919, App	219	19	100.0	225	12	US-10-424-599-262650	Sequence 262650, App
147	19	100.0	178	9	US-09-925-301-914	Sequence 914, App	220	19	100.0	225	12	US-10-351-334-155	Sequence 155, App
148	19	100.0	178	9	US-09-925-297-592	Sequence 592, App	221	19	100.0	225	12	US-10-156-761-14916	Sequence 14916, A
149	19	100.0	178	15	US-10-264-049-5578	Sequence 5578, App	222	19	100.0	225	12	US-10-369-493-14363	Sequence 14363, App
150	19	100.0	180	12	US-10-424-599-170244	Sequence 170244, App	223	19	100.0	225	15	US-10-369-493-14363	Sequence 4027, App
151	19	100.0	180	14	US-10-156-761-12228	Sequence 12228, App	224	19	100.0	226	9	US-09-738-626-4027	Sequence 53174, A
152	19	100.0	184	12	US-10-335-977-7279	Sequence 7279, App	225	19	100.0	226	12	US-10-425-114-53174	Sequence 53174, A
153	19	100.0	187	9	US-09-804-156-17	Sequence 17, Appl	226	19	100.0	226	9	US-09-738-626-4027	Sequence 53174, A
154	19	100.0	187	9	US-09-946-633-9	Sequence 9, Appl	227	19	100.0	226	9	US-09-738-626-4027	Sequence 53174, A
155	19	100.0	187	13	US-10-125-459-9	Sequence 9, Appl	228	19	100.0	227	16	US-10-275-505-7	Sequence 7, Appli
156	19	100.0	187	13	US-10-067-761-17	Sequence 17, Appl	229	19	100.0	228	12	US-10-425-114-47426	Sequence 47426, A
157	19	100.0	187	14	US-10-319-519-17	Sequence 17, Appl	230	19	100.0	228	13	US-10-051-325-8	Sequence 8, Appli
158	19	100.0	188	9	US-09-734-563-134	Sequence 134, App	231	19	100.0	229	12	US-10-424-599-143683	Sequence 143683, App
159	19	100.0	188	12	US-10-282-122A-76866	Sequence 76866, A	232	19	100.0	230	9	US-09-738-626-4027	Sequence 53174, A
160	19	100.0	188	12	US-10-282-122A-78177	Sequence 78177, A	233	19	100.0	231	12	US-10-425-114-53174	Sequence 53174, A
161	19	100.0	188	12	US-10-424-599-256741	Sequence 256741, App	234	19	100.0	231	12	US-10-425-114-53174	Sequence 53174, A

235	19	100.0	232	10	US-09-252-088-40	Sequence 40, Appl	308	19	100.0	270	9	US-09-738-626-3943	Sequence 3943, Ap
236	19	100.0	232	15	US-10-340-792-40	Sequence 40, Appl	309	19	100.0	270	12	US-10-262-511-148	Sequence 148, Appl
237	19	100.0	233	12	US-10-425-114-45098	Sequence 45098, A	310	19	100.0	271	15	US-10-369-493-10566	Sequence 10566, A
238	19	100.0	234	12	US-10-425-114-63330	Sequence 63330, A	311	19	100.0	271	15	US-10-264-237-3658	Sequence 2658, A
239	19	100.0	234	12	US-10-425-114-64490	Sequence 64490, A	312	19	100.0	272	10	US-09-764-891-3960	Sequence 2960, Ap
240	19	100.0	235	9	US-09-815-242-11738	Sequence 11738, A	313	19	100.0	272	12	US-10-282-132A-70355	Sequence 70355, A
241	19	100.0	236	12	US-10-412-6998-1410	Sequence 1410, Ap	314	19	100.0	273	14	US-10-182-447-3	Sequence 3, Appl
242	19	100.0	236	15	US-10-421-138A-308	Sequence 308, Appl	315	19	100.0	274	12	US-10-282-122A-51559	Sequence 51559, A
243	19	100.0	236	15	US-10-374-780A-1221	Sequence 1221, Ap	316	19	100.0	274	12	US-10-282-122A-74371	Sequence 74371, A
244	19	100.0	238	12	US-10-282-122A-70662	Sequence 70662, A	317	19	100.0	275	12	US-10-282-122A-52045	Sequence 52045, A
245	19	100.0	238	12	US-10-282-122A-72399	Sequence 72399, A	318	19	100.0	275	12	US-10-424-599-146071	Sequence 146071, A
246	19	100.0	238	14	US-10-137-036-68	Sequence 68, Appl	319	19	100.0	278	12	US-10-282-122A-65461	Sequence 65461, A
247	19	100.0	239	12	US-10-424-599-257488	Sequence 257488, A	320	19	100.0	278	12	US-10-282-122A-65761	Sequence 65761, A
248	19	100.0	240	14	US-10-338-777-39	Sequence 39, Appl	321	19	100.0	279	9	US-10-369-493-11888	Sequence 11888, A
249	19	100.0	241	12	US-10-282-122A-56234	Sequence 56234, A	322	19	100.0	280	9	US-09-738-626-4505	Sequence 4505, Ap
250	19	100.0	241	12	US-10-425-114-36715	Sequence 36715, A	323	19	100.0	280	15	US-10-369-493-20769	Sequence 20769, A
251	19	100.0	242	12	US-10-425-114-47756	Sequence 47756, A	324	19	100.0	281	13	US-10-059-964-58	Sequence 58, Appl
252	19	100.0	243	10	US-09-870-406A-44	Sequence 44, Appl	325	19	100.0	281	14	US-10-314-639-58	Sequence 58, Appl
253	19	100.0	243	10	US-10-424-599-257488	Sequence 257488, A	326	19	100.0	281	14	US-10-156-761-9102	Sequence 9102, Ap
254	19	100.0	243	14	US-10-159-901-44	Sequence 44, Appl	327	19	100.0	282	9	US-09-925-300-1694	Sequence 1694, Ap
255	19	100.0	244	12	US-10-282-122A-72187	Sequence 72187, A	328	19	100.0	282	12	US-10-282-122A-47275	Sequence 47275, A
256	19	100.0	244	12	US-10-424-599-158126	Sequence 158126, A	329	19	100.0	282	14	US-10-156-761-12355	Sequence 12355, A
257	19	100.0	244	12	US-10-424-599-271081	Sequence 271081, A	330	19	100.0	283	8	US-08-781-986A-5203	Sequence 5203, Ap
258	19	100.0	246	9	US-09-771-161A-129	Sequence 129, Appl	331	19	100.0	283	12	US-10-425-114-43988	Sequence 43988, A
259	19	100.0	246	14	US-10-156-761-13663	Sequence 13663, A	332	19	100.0	283	12	US-10-323-624-5203	Sequence 5203, Ap
260	19	100.0	245	15	US-10-421-138A-311	Sequence 311, Appl	333	19	100.0	284	12	US-10-425-114-58685	Sequence 58685, A
261	19	100.0	245	15	US-10-374-780A-1244	Sequence 1244, Ap	334	19	100.0	284	12	US-10-282-132A-49530	Sequence 49530, A
262	19	100.0	248	12	US-10-425-114-4065	Sequence 4065, A	335	19	100.0	285	12	US-10-282-132A-69922	Sequence 69922, A
263	19	100.0	249	9	US-09-822-312-9	Sequence 9, Appl	336	19	100.0	286	12	US-10-425-114-64235	Sequence 64235, A
264	19	100.0	249	11	US-09-829-495-9	Sequence 9, Appl	337	19	100.0	286	12	US-10-282-122A-69316	Sequence 69316, A
265	19	100.0	249	15	US-10-156-761-9204	Sequence 9204, Ap	338	19	100.0	287	12	US-10-262-511-1471	Sequence 147161, A
266	19	100.0	249	15	US-10-409-701-15	Sequence 15, Appl	339	19	100.0	287	12	US-10-424-599-144161	Sequence 20884, A
267	19	100.0	249	15	US-10-310-154-606	Sequence 606, Appl	340	19	100.0	287	15	US-10-369-493-20884	Sequence 66912, A
268	19	100.0	253	9	US-09-835-242-10279	Sequence 10279, A	341	19	100.0	288	12	US-10-282-132A-66912	Sequence 66912, A
269	19	100.0	253	15	US-10-369-493-25327	Sequence 25327, A	342	19	100.0	288	9	US-09-815-242-11193	Sequence 66363, A
270	19	100.0	254	12	US-10-424-599-158554	Sequence 158554, A	343	19	100.0	288	12	US-10-425-114-66363	Sequence 11193, A
271	19	100.0	254	12	US-10-424-599-198953	Sequence 198953, A	344	19	100.0	289	12	US-10-282-122A-58429	Sequence 58429, A
272	19	100.0	254	15	US-10-362-511-150	Sequence 150, Appl	345	19	100.0	289	12	US-10-425-114-47196	Sequence 47196, A
273	19	100.0	254	15	US-10-369-493-11429	Sequence 11429, A	346	19	100.0	290	12	US-10-424-599-196081	Sequence 196081, A
274	19	100.0	255	14	US-10-086-623-40	Sequence 40, Appl	347	19	100.0	290	12	US-10-262-511-146	Sequence 146, Appl
275	19	100.0	255	14	US-10-360-539-40	Sequence 40, Appl	348	19	100.0	290	15	US-10-369-493-15865	Sequence 15865, A
276	19	100.0	255	15	US-10-369-493-19054	Sequence 19054, A	349	19	100.0	291	12	US-10-041-018-27	Sequence 27, Appl
277	19	100.0	256	15	US-10-421-138A-312	Sequence 312, Appl	350	19	100.0	291	14	US-10-017-161-2430	Sequence 2430, Ap
278	19	100.0	256	15	US-10-374-780A-1247	Sequence 1247, Ap	351	19	100.0	291	15	US-10-292-798-2070	Sequence 2070, Ap
279	19	100.0	257	9	US-09-738-626-6642	Sequence 6642, Ap	352	19	100.0	292	15	US-10-264-237-2770	Sequence 2770, Ap
280	19	100.0	257	14	US-10-106-698-5221	Sequence 5221, Ap	353	19	100.0	292	15	US-10-369-493-5341	Sequence 5341, Ap
281	19	100.0	257	15	US-10-139-794-16	Sequence 16, Appl	354	19	100.0	292	15	US-10-369-493-15494	Sequence 15494, A
282	19	100.0	258	9	US-09-815-242-11889	Sequence 11889, A	355	19	100.0	293	12	US-10-282-132A-76466	Sequence 76466, A
283	19	100.0	258	12	US-10-282-122A-66426	Sequence 66426, A	356	19	100.0	293	14	US-10-156-761-14726	Sequence 14726, A
284	19	100.0	258	12	US-10-424-599-248969	Sequence 248969, A	357	19	100.0	294	12	US-10-299-867-20	Sequence 20, Appl
285	19	100.0	258	14	US-10-097-111-303	Sequence 303, Appl	358	19	100.0	294	14	US-10-236-699-12	Sequence 12, Appl
286	19	100.0	260	12	US-10-425-114-46566	Sequence 46566, A	359	19	100.0	294	15	US-10-306-762-155	Sequence 155, Appl
287	19	100.0	261	10	US-09-893-519A-19	Sequence 19, Appl	360	19	100.0	294	15	US-10-369-493-13821	Sequence 13821, A
288	19	100.0	261	12	US-10-282-122A-69377	Sequence 69377, A	361	19	100.0	295	14	US-10-306-762-141	Sequence 141, Appl
289	19	100.0	261	14	US-10-011-364-10	Sequence 10, Appl	362	19	100.0	297	12	US-10-282-132A-50243	Sequence 50243, A
290	19	100.0	261	15	US-10-369-493-12238	Sequence 12238, A	363	19	100.0	297	12	US-10-282-132A-60194	Sequence 60194, A
291	19	100.0	261	15	US-10-321-562-10	Sequence 10, Appl	364	19	100.0	297	12	US-10-425-114-41342	Sequence 41342, A
292	19	100.0	262	12	US-10-424-599-254101	Sequence 254101, A	365	19	100.0	299	9	US-09-393-634-35	Sequence 35, Appl
293	19	100.0	263	9	US-09-800-729-88	Sequence 88, Appl	366	19	100.0	299	10	US-09-510-332-1	Sequence 1, Appl
294	19	100.0	263	14	US-09-746-783-2	Sequence 2, Appl	367	19	100.0	299	10	US-09-832-522-73	Sequence 73, Appl
295	19	100.0	263	11	US-09-633-245-2213	Sequence 2213, Ap	368	19	100.0	299	12	US-10-364-861-35	Sequence 35, Appl
296	19	100.0	264	14	US-10-182-447-1	Sequence 1, Appl	369	19	100.0	299	13	US-10-059-964-46	Sequence 46, Appl
297	19	100.0	265	12	US-10-282-122A-54414	Sequence 54414, A	370	19	100.0	299	14	US-10-314-639-46	Sequence 46, Appl
298	19	100.0	265	12	US-10-425-114-41437	Sequence 41437, A	371	19	100.0	299	14	US-10-383-982-35	Sequence 35, Appl
299	19	100.0	265	12	US-10-425-114-62005	Sequence 62005, A	372	19	100.0	303	9	US-09-738-626-3734	Sequence 3734, Ap
300	19	100.0	266	12	US-10-424-599-252581	Sequence 252581, A	373	19	100.0	304	12	US-10-425-114-53342	Sequence 53342, A
301	19	100.0	266	12	US-10-282-122A-78178	Sequence 78178, A	374	19	100.0	304	12	US-10-424-599-219921	Sequence 219921, A
302	19	100.0	267	12	US-10-425-114-45104	Sequence 45104, A	375	19	100.0	304	12	US-10-425-114-62081	Sequence 62081, A
303	19	100.0	267	12	US-10-425-114-66738	Sequence 66738, A	376	19	100.0	305	9	US-10-369-493-9667	Sequence 9667, Ap
304	19	100.0	267	15	US-10-421-138A-315	Sequence 315, Appl	377	19	100.0	306	12	US-09-738-626-6591	Sequence 6591, Ap
305	19	100.0	268	9	US-09-745-763-164	Sequence 164, Appl	378	19	100.0	306	12	US-10-425-114-62473	Sequence 62473, A
306	19	100.0	268	16	US-10-358-662-19	Sequence 19, Appl	379	19	100.0	306	12	US-10-425-114-64440	Sequence 64440, A
307	19	100.0	269	15	US-10-104-047-2182	Sequence 2182, Ap	380	19	100.0	306	12	US-10-425-114-67258	Sequence 67258, A

381	19	100.0	306	14	US-10-091-007-40	Sequence 40, Appl	454	19	100.0	321	14	US-10-025-806-32	Sequence 32, Appl
382	19	100.0	306	15	US-10-320-797-3084	Sequence 3084, Ap	455	19	100.0	322	12	US-10-282-122A-53145	Sequence 53145, A
383	19	100.0	307	14	US-10-156-761-8528	Sequence 8528, Ap	456	19	100.0	322	14	US-10-270-861-7	Sequence 7, Appl
384	19	100.0	308	9	US-09-925-299-930	Sequence 930, App	457	19	100.0	322	14	US-10-270-861-11	Sequence 11, Appl
385	19	100.0	308	9	US-09-738-626-6728	Sequence 6728, Ap	458	19	100.0	323	12	US-10-282-122A-69306	Sequence 69306, A
386	19	100.0	308	10	US-09-928-299-930	Sequence 930, App	459	19	100.0	323	12	US-10-424-599-238938	Sequence 238938, A
387	19	100.0	308	14	US-10-032-201B-253	Sequence 253, App	460	19	100.0	323	12	US-10-282-122A-57182	Sequence 57182, A
388	19	100.0	309	12	US-10-423-114-45522	Sequence 45522, A	461	19	100.0	324	12	US-10-282-122A-61308	Sequence 61308, A
389	19	100.0	309	14	US-10-023-601-36	Sequence 36, Appl	462	19	100.0	324	14	US-10-156-761-13399	Sequence 13399, A
390	19	100.0	309	15	US-10-363-493-7640	Sequence 7640, Ap	463	19	100.0	324	14	US-10-156-761-14083	Sequence 14083, A
391	19	100.0	309	15	US-10-363-493-11118	Sequence 11118, A	464	19	100.0	325	12	US-10-262-511-152	Sequence 152, App
392	19	100.0	310	9	US-09-826-909-2	Sequence 2, Appli	465	19	100.0	325	14	US-10-270-861-1	Sequence 1, Appli
393	19	100.0	310	12	US-10-282-122A-67405	Sequence 67405, A	466	19	100.0	325	14	US-10-270-861-13	Sequence 13, Appli
394	19	100.0	310	12	US-10-424-599-163497	Sequence 163497, A	467	19	100.0	326	15	US-10-369-493-1001	Sequence 1001, Ap
395	19	100.0	310	12	US-10-424-599-230916	Sequence 230916, A	468	19	100.0	326	15	US-10-424-599-214738	Sequence 214738, A
396	19	100.0	310	15	US-10-191-121-18	Sequence 18, Appl	469	19	100.0	327	12	US-10-425-114-52852	Sequence 52852, A
397	19	100.0	310	16	US-09-886-0534-16	Sequence 16, Appl	470	19	100.0	327	14	US-10-314-657-8	Sequence 8, Appli
398	19	100.0	311	9	US-09-886-055-9	Sequence 9, Appli	471	19	100.0	328	9	US-09-815-242-10935	Sequence 10935, A
399	19	100.0	311	9	US-09-738-626-3671	Sequence 3671, Ap	472	19	100.0	328	10	US-09-765-061B-75	Sequence 75, Appl
400	19	100.0	311	10	US-09-804-291-9	Sequence 9, Appli	473	19	100.0	328	10	US-09-765-061B-76	Sequence 76, Appl
401	19	100.0	311	10	US-09-833-522-20	Sequence 20, Appl	474	19	100.0	328	12	US-10-282-122A-56878	Sequence 56878, A
402	19	100.0	311	10	US-09-795-271-25	Sequence 25, Appl	475	19	100.0	328	15	US-10-369-493-20838	Sequence 20838, A
403	19	100.0	311	11	US-09-981-151A-75	Sequence 75, Appl	476	19	100.0	330	9	US-09-771-161A-128	Sequence 128, App
404	19	100.0	311	11	US-09-844-861A-12	Sequence 12, Appl	477	19	100.0	331	9	US-09-815-242-11761	Sequence 11761, A
405	19	100.0	311	12	US-10-343-650A-260	Sequence 260, App	478	19	100.0	331	12	US-10-424-599-176898	Sequence 176898, A
406	19	100.0	311	12	US-10-343-650A-364	Sequence 364, App	479	19	100.0	331	12	US-10-425-114-43781	Sequence 43781, A
407	19	100.0	311	14	US-10-023-597-52	Sequence 52, Appl	480	19	100.0	334	12	US-10-382-122A-71145	Sequence 71145, A
408	19	100.0	311	14	US-10-017-161-280	Sequence 280, App	481	19	100.0	334	15	US-10-264-049-2446	Sequence 2446, Ap
409	19	100.0	311	14	US-10-017-161-284	Sequence 284, App	482	19	100.0	335	12	US-09-984-271-118	Sequence 118, App
410	19	100.0	311	15	US-10-008-041A-20	Sequence 20, Appl	483	19	100.0	335	12	US-10-425-114-39152	Sequence 39152, A
411	19	100.0	311	15	US-10-293-798-250	Sequence 250, App	484	19	100.0	335	12	US-09-984-276-118	Sequence 118, App
412	19	100.0	311	15	US-10-293-798-254	Sequence 254, App	485	19	100.0	335	15	US-10-369-493-1192	Sequence 1192, A
413	19	100.0	312	9	US-09-886-055-187	Sequence 187, App	486	19	100.0	336	12	US-10-282-122A-67767	Sequence 67767, A
414	19	100.0	312	9	US-09-738-626-5007	Sequence 5007, Ap	487	19	100.0	336	12	US-10-282-122A-78253	Sequence 78253, A
415	19	100.0	312	10	US-09-804-291-187	Sequence 187, App	488	19	100.0	336	12	US-10-425-114-46555	Sequence 46555, A
416	19	100.0	312	10	US-10-182-122A-53997	Sequence 53997, A	489	19	100.0	336	15	US-10-369-493-419	Sequence 419, App
417	19	100.0	312	12	US-10-282-122A-17	Sequence 17, Appl	490	19	100.0	337	10	US-09-933-705-42	Sequence 42, Appl
418	19	100.0	312	12	US-10-231-079-34	Sequence 34, Appl	491	19	100.0	337	12	US-10-425-114-57499	Sequence 57499, A
419	19	100.0	312	12	US-10-343-650A-656	Sequence 656, App	492	19	100.0	337	12	US-10-425-114-59818	Sequence 59818, A
420	19	100.0	312	14	US-10-017-161-50	Sequence 50, Appl	493	19	100.0	337	12	US-10-425-114-63730	Sequence 63730, A
421	19	100.0	312	14	US-10-017-161-498	Sequence 498, App	494	19	100.0	337	15	US-10-369-493-11195	Sequence 11195, A
422	19	100.0	312	15	US-10-387-629-16	Sequence 16, Appl	495	19	100.0	338	9	US-09-852-976-1	Sequence 1, Appli
423	19	100.0	312	15	US-10-293-798-42	Sequence 42, Appl	496	19	100.0	338	12	US-10-425-114-53202	Sequence 53202, A
424	19	100.0	313	15	US-10-369-493-11089	Sequence 11089, A	497	19	100.0	338	12	US-10-425-114-61153	Sequence 61153, A
425	19	100.0	314	12	US-10-282-122A-45385	Sequence 45385, A	498	19	100.0	338	12	US-10-425-114-62150	Sequence 62150, A
426	19	100.0	314	12	US-10-424-599-181578	Sequence 181578, A	499	19	100.0	338	12	US-10-425-114-66618	Sequence 66618, A
427	19	100.0	314	12	US-10-424-599-185947	Sequence 185947, A	500	19	100.0	339	9	US-09-832-312-34	Sequence 34, Appli
428	19	100.0	315	12	US-10-424-599-189944	Sequence 189944, A	501	19	100.0	339	9	US-09-832-312-36	Sequence 36, Appl
429	19	100.0	315	14	US-10-023-597-102	Sequence 102, App	502	19	100.0	339	9	US-09-832-312-38	Sequence 38, Appl
430	19	100.0	315	14	US-10-238-075-812	Sequence 812, App	503	19	100.0	339	9	US-09-832-312-40	Sequence 40, Appl
431	19	100.0	315	14	US-10-311-626-1	Sequence 1, Appli	504	19	100.0	339	11	US-09-829-495-3	Sequence 3, Appli
432	19	100.0	315	16	US-10-309-290-138	Sequence 138, App	505	19	100.0	339	11	US-09-829-495-34	Sequence 34, Appl
433	19	100.0	316	12	US-10-072-012-407	Sequence 407, App	506	19	100.0	339	11	US-09-829-495-36	Sequence 36, Appl
434	19	100.0	316	14	US-10-236-699-16	Sequence 16, Appl	507	19	100.0	339	11	US-09-829-495-38	Sequence 38, Appl
435	19	100.0	317	12	US-10-425-114-43447	Sequence 43447, A	508	19	100.0	339	11	US-09-829-495-40	Sequence 40, Appl
436	19	100.0	318	10	US-09-795-271-40	Sequence 40, Appl	509	19	100.0	339	11	US-09-829-495-40	Sequence 40, Appl
437	19	100.0	318	10	US-09-907-218-63	Sequence 63, Appl	510	19	100.0	339	14	US-10-157-031-387	Sequence 387, App
438	19	100.0	318	11	US-09-965-422-41	Sequence 41, Appl	511	19	100.0	339	14	US-10-156-761-14680	Sequence 14680, A
439	19	100.0	318	11	US-09-981-566A-42	Sequence 42, Appl	512	19	100.0	339	14	US-10-053-248-34	Sequence 34, Appl
440	19	100.0	318	11	US-09-981-566A-47	Sequence 47, Appl	513	19	100.0	339	14	US-10-446-826-5	Sequence 5, Appli
441	19	100.0	318	14	US-10-032-201B-284	Sequence 284, App	514	19	100.0	342	12	US-10-425-114-48872	Sequence 48872, A
442	19	100.0	318	14	US-10-375-611-1	Sequence 1, Appli	515	19	100.0	342	12	US-10-276-774-2289	Sequence 2289, Ap
443	19	100.0	318	14	US-10-247-671-131	Sequence 131, App	516	19	100.0	342	12	US-09-805-020-49	Sequence 49, Appl
444	19	100.0	319	9	US-09-832-312-5	Sequence 5, Appli	517	19	100.0	342	14	US-10-238-075-285	Sequence 285, App
445	19	100.0	319	11	US-09-828-495-5	Sequence 5, Appli	518	19	100.0	345	9	US-09-935-300-1188	Sequence 1188, Ap
446	19	100.0	319	12	US-10-425-114-70828	Sequence 70828, A	519	19	100.0	345	12	US-10-424-599-223433	Sequence 223433
447	19	100.0	319	13	US-10-001-426-4	Sequence 4, Appli	520	19	100.0	346	10	US-09-953-348-107	Sequence 107, App
448	19	100.0	319	13	US-10-003-014-4	Sequence 4, Appli	521	19	100.0	346	12	US-10-425-114-66218	Sequence 66218, A
449	19	100.0	319	14	US-10-236-699-10	Sequence 10, Appl	522	19	100.0	346	12	US-10-425-114-67904	Sequence 67904, A
450	19	100.0	320	12	US-10-424-599-186648	Sequence 186648, A	523	19	100.0	346	14	US-10-267-255-107	Sequence 107, App
451	19	100.0	320	12	US-10-335-977-5813	Sequence 5813, Ap	524	19	100.0	347	9	US-09-738-626-6922	Sequence 6922, Ap
452	19	100.0	320	16	US-10-461-990-28	Sequence 28, Appl	525	19	100.0	347	12	US-10-425-114-52822	Sequence 52822, A
453	19	100.0	321	12	US-10-424-599-163498	Sequence 163498, A	526	19	100.0	347	14	US-10-306-762-240	Sequence 240, App

527 19 100.0 347 15 US-10-264-049-2882 Sequence 2882, Ap  
528 19 100.0 349 12 US-10-335-977-5814 Sequence 5814, Ap  
529 19 100.0 349 16 US-10-389-566-1121 Sequence 1121, Ap  
530 19 100.0 349 16 US-10-389-566-1121 Sequence 1121, Ap  
531 19 100.0 351 12 US-10-282-122A-46888 Sequence 46888, A  
532 19 100.0 351 15 US-10-369-493-10859 Sequence 10859, A  
533 19 100.0 351 15 US-10-369-493-10888 Sequence 10888, A  
534 19 100.0 352 12 US-10-282-122A-60673 Sequence 60673, A  
535 19 100.0 352 16 US-10-389-566-20240 Sequence 20240, A  
536 19 100.0 352 16 US-10-389-566-2165 Sequence 2165, Ap  
537 19 100.0 353 14 US-10-370-861-9 Sequence 9, Appl  
538 19 100.0 354 16 US-10-616-624-4 Sequence 4, Appl  
539 19 100.0 355 14 US-10-106-698-4846 Sequence 4846, Ap  
540 19 100.0 357 12 US-10-363-616-438 Sequence 438, Ap  
541 19 100.0 358 12 US-10-087-192-1845 Sequence 1845, Ap  
542 19 100.0 359 14 US-10-032-585-7893 Sequence 7893, A  
543 19 100.0 360 12 US-10-425-114-65637 Sequence 65637, A  
544 19 100.0 363 12 US-10-425-114-61874 Sequence 61874, A  
545 19 100.0 364 12 US-10-424-599-214841 Sequence 214841, A  
546 19 100.0 364 13 US-10-086-623-38 Sequence 38, Appl  
547 19 100.0 364 14 US-10-260-539-38 Sequence 38, Appl  
548 19 100.0 365 14 US-10-156-761-8947 Sequence 8947, Ap  
549 19 100.0 367 12 US-10-425-114-64842 Sequence 64842, A  
550 19 100.0 368 12 US-10-424-599-227208 Sequence 227208, A  
551 19 100.0 368 15 US-10-369-493-145 Sequence 145, Ap  
552 19 100.0 368 15 US-10-369-493-2742 Sequence 2742, Ap  
553 19 100.0 369 12 US-10-282-122A-56898 Sequence 56898, A  
554 19 100.0 369 12 US-10-282-122A-73956 Sequence 73956, A  
555 19 100.0 369 12 US-10-282-122A-74479 Sequence 74479, A  
556 19 100.0 369 14 US-10-115-571A-31 Sequence 31, Appl  
557 19 100.0 369 14 US-10-446-826-37 Sequence 37, Appl  
558 19 100.0 370 9 US-09-808-972-4 Sequence 4, Appl  
559 19 100.0 370 11 US-09-876-813-53 Sequence 53, Appl  
560 19 100.0 370 12 US-10-606-055-4 Sequence 4, Appl  
561 19 100.0 370 12 US-10-365-095-4 Sequence 4, Appl  
562 19 100.0 370 12 US-10-086-623-36 Sequence 36, Appl  
563 19 100.0 370 13 US-10-039-847A-4 Sequence 4, Appl  
564 19 100.0 370 14 US-10-260-539-36 Sequence 36, Appl  
565 19 100.0 370 14 US-10-226-559-4 Sequence 4, Appl  
566 19 100.0 370 14 US-10-274-638-4 Sequence 4, Appl  
567 19 100.0 370 14 US-10-011-364-8 Sequence 8, Appl  
568 19 100.0 371 15 US-10-321-962-8 Sequence 8, Appl  
569 19 100.0 371 15 US-10-282-122A-72020 Sequence 72020, A  
570 19 100.0 371 12 US-10-282-122A-73684 Sequence 73684, A  
571 19 100.0 371 12 US-10-282-122A-74303 Sequence 74303, A  
572 19 100.0 371 12 US-10-424-599-218495 Sequence 218495, A  
573 19 100.0 372 9 US-09-800-729-213 Sequence 213, Appl  
574 19 100.0 372 10 US-09-765-061B-74 Sequence 74, Appl  
575 19 100.0 372 10 US-09-765-061B-78 Sequence 78, Appl  
576 19 100.0 374 9 US-09-815-242-13447 Sequence 13447, A  
577 19 100.0 374 9 US-09-815-242-13692 Sequence 13692, A  
578 19 100.0 374 12 US-10-425-114-38772 Sequence 38772, A  
579 19 100.0 374 14 US-10-029-386-33411 Sequence 33411, A  
580 19 100.0 375 9 US-09-815-242-10288 Sequence 10288, A  
581 19 100.0 377 12 US-10-282-122A-53791 Sequence 53791, A  
582 19 100.0 378 9 US-09-971-536-46 Sequence 46, Appl  
583 19 100.0 378 15 US-10-369-493-13902 Sequence 13902, A  
584 19 100.0 379 12 US-10-282-122A-51871 Sequence 51871, A  
585 19 100.0 379 12 US-10-282-122A-77140 Sequence 77140, A  
586 19 100.0 379 14 US-10-205-331-68 Sequence 68, Appl  
587 19 100.0 379 14 US-10-157-223-6 Sequence 6, Appl  
588 19 100.0 380 9 US-09-815-242-11291 Sequence 11291, A  
589 19 100.0 380 12 US-09-882-227-466 Sequence 466, Ap  
590 19 100.0 380 12 US-10-282-122A-61191 Sequence 61191, A  
591 19 100.0 381 12 US-10-425-114-65079 Sequence 65079, A  
592 19 100.0 381 12 US-10-282-122A-61883 Sequence 61883, A  
593 19 100.0 382 15 US-10-369-493-8252 Sequence 8252, Ap  
594 19 100.0 383 12 US-10-282-122A-43909 Sequence 43909, A  
595 19 100.0 383 14 US-10-166-225A-76 Sequence 76, Appl  
596 19 100.0 383 16 US-10-389-566-1853 Sequence 1853, Ap  
597 19 100.0 384 10 US-09-765-061B-72 Sequence 72, Appl  
598 19 100.0 384 10 US-09-765-061B-73 Sequence 73, Appl  
599 19 100.0 384 14 US-10-156-761-14937 Sequence 14937, A

Sequence 18459, A  
Sequence 67317, A  
Sequence 6719, Ap  
Sequence 11041, A  
Sequence 58201, A  
Sequence 4047, Ap  
Sequence 264035, A  
Sequence 59654, A  
Sequence 82, Appl  
Sequence 83, Appl  
Sequence 20787, A  
Sequence 8879, Ap  
Sequence 18574, A  
Sequence 3160, Ap  
Sequence 77, Appl  
Sequence 50194, A  
Sequence 11600, A  
Sequence 14301, A  
Sequence 14467, A  
Sequence 15069, A  
Sequence 19323, Ap  
Sequence 4, Appl  
Sequence 64, Appl  
Sequence 8, Appl  
Sequence 54557, A  
Sequence 50770, A  
Sequence 4, Appl  
Sequence 103, Appl  
Sequence 121, Appl  
Sequence 2612, Ap  
Sequence 8, Appl  
Sequence 236415, A  
Sequence 88, Appl  
Sequence 152, Appl  
Sequence 47722, A  
Sequence 74248, A  
Sequence 18573, A  
Sequence 74, Appl  
Sequence 212357, A  
Sequence 1926, Ap  
Sequence 225710, A  
Sequence 55610, A  
Sequence 72509, A  
Sequence 21706, A  
Sequence 68690, A  
Sequence 19613, A  
Sequence 635, Appl  
Sequence 63, Appl  
Sequence 13889, A  
Sequence 57253, A  
Sequence 438, Ap  
Sequence 438, Ap  
Sequence 74832, A  
Sequence 193438, A  
Sequence 8848, Ap  
Sequence 95, Appl  
Sequence 95, Appl  
Sequence 95, Appl  
Sequence 5136, Ap  
Sequence 9322, Ap  
Sequence 14, Appl  
Sequence 266089, A  
Sequence 23, Appl  
Sequence 78449, A  
Sequence 1669, Ap  
Sequence 3, Appl  
Sequence 42865, A  
Sequence 54669, A  
Sequence 680, Ap  
Sequence 3, Appl  
Sequence 680, Ap  
Sequence 7896, Ap

US-10-369-493-18459  
US-10-425-114-67317  
US-10-369-493-6719  
US-09-815-242-11041  
US-10-282-122A-58201  
US-10-369-493-4047  
US-10-424-599-284035  
US-10-425-114-59654  
US-10-166-225A-82  
US-10-166-225A-83  
US-10-369-493-20787  
US-10-369-493-8879  
US-10-369-493-18574  
US-10-369-493-3160  
US-09-765-061B-77  
US-10-425-114-50194  
US-10-369-493-11600  
US-10-369-493-14301  
US-10-369-493-14467  
US-10-369-493-15069  
US-10-087-192-1923  
US-09-334-903-4  
US-09-334-868-64  
US-09-941-947A-8  
US-10-282-122A-54557  
US-10-425-114-50770  
US-10-700-003-4  
US-10-381-779-103  
US-10-381-779-121  
US-10-264-049-2612  
US-10-363-567-8  
US-10-424-599-236415  
US-10-204-887-88  
US-09-769-744A-152  
US-10-282-122A-47722  
US-10-282-122A-74248  
US-10-369-493-18573  
US-10-166-225A-74  
US-10-424-599-212357  
US-10-087-192-1926  
US-10-424-599-225710  
US-10-282-122A-55610  
US-10-282-122A-72509  
US-10-369-493-21706  
US-10-282-122A-68690  
US-10-369-493-19613  
US-10-389-647-635  
US-10-403-877-63  
US-10-363-493-13889  
US-10-282-122A-57253  
US-10-287-226-436  
US-10-287-226-438  
US-10-282-122A-74832  
US-10-424-599-193438  
US-10-156-761-8848  
US-10-024-218A-95  
US-10-617-217A-95  
US-09-738-626-5136  
US-10-156-761-9322  
US-10-081-816-14  
US-10-424-599-266089  
US-10-081-816-23  
US-10-282-122A-78449  
US-09-925-300-1669  
US-09-300-425A-3  
US-10-282-122A-42865  
US-10-425-114-54669  
US-09-925-298-680  
US-10-073-185-3  
US-10-102-806-680  
US-10-156-761-7850  
US-10-156-761-7896

673 19 100.0 413 12 US-10-282-132A-48732 Sequence 48732, A 746  
674 19 100.0 413 15 US-10-369-493-1140 Sequence 1140, Ap  
675 19 100.0 414 9 US-09-820-893-59 Sequence 69, Appl  
676 19 100.0 414 12 US-10-424-599-285027 Sequence 285027, A  
677 19 100.0 414 12 US-10-607-565-69 Sequence 69, Appl  
678 19 100.0 414 16 US-10-287-226-440 Sequence 440, Appl  
679 19 100.0 415 12 US-10-282-132A-65349 Sequence 65349, A  
680 19 100.0 415 12 US-10-282-132A-65787 Sequence 65787, A  
681 19 100.0 415 12 US-10-282-132A-68742 Sequence 68742, A  
682 19 100.0 416 14 US-10-156-761-12502 Sequence 12502, A  
683 19 100.0 416 15 US-10-369-493-4737 Sequence 4737, Ap  
684 19 100.0 416 15 US-10-369-493-4882 Sequence 4882, Ap  
685 19 100.0 416 15 US-10-369-493-7496 Sequence 7496, Ap  
686 19 100.0 417 12 US-10-424-599-274539 Sequence 274539, A  
687 19 100.0 417 12 US-10-425-114-56059 Sequence 56059, A  
688 19 100.0 417 14 US-10-128-714-3345 Sequence 3345, Ap  
689 19 100.0 417 14 US-10-128-714-3345 Sequence 3345, Ap  
690 19 100.0 418 11 US-09-981-151A-29 Sequence 29, Appl  
691 19 100.0 419 15 US-10-369-493-20475 Sequence 20475, A  
692 19 100.0 420 15 US-10-369-493-13961 Sequence 13961, A  
693 19 100.0 421 9 US-09-815-243-5303 Sequence 5303, Ap  
694 19 100.0 421 9 US-09-815-243-13388 Sequence 13388, A  
695 19 100.0 421 9 US-09-815-243-12820 Sequence 12820, A  
696 19 100.0 421 12 US-10-282-132A-71223 Sequence 71223, A  
697 19 100.0 421 12 US-10-425-114-56431 Sequence 56431, A  
698 19 100.0 421 14 US-10-138-701-8 Sequence 8, Appl  
699 19 100.0 421 14 US-10-138-701-22 Sequence 22, Appl  
700 19 100.0 422 12 US-10-282-132A-66702 Sequence 66702, A  
701 19 100.0 423 12 US-10-282-132A-53283 Sequence 53283, A  
702 19 100.0 423 12 US-10-424-599-204268 Sequence 204268, A  
703 19 100.0 423 12 US-10-425-114-46740 Sequence 46740, A  
704 19 100.0 423 12 US-10-107-431-49 Sequence 49, Appl  
705 19 100.0 423 12 US-10-369-493-222 Sequence 222, Appl  
706 19 100.0 423 15 US-10-369-493-5557 Sequence 5557, Ap  
707 19 100.0 424 12 US-10-282-132A-45014 Sequence 45014, A  
708 19 100.0 424 12 US-10-424-599-214840 Sequence 214840, A  
709 19 100.0 424 14 US-10-156-761-14870 Sequence 14870, A  
710 19 100.0 425 14 US-10-156-761-7776 Sequence 7776, Ap  
711 19 100.0 425 14 US-10-214-446-20 Sequence 20, Appl  
712 19 100.0 426 12 US-10-425-114-48506 Sequence 48506, A  
713 19 100.0 427 15 US-10-369-493-2977 Sequence 2977, Ap  
714 19 100.0 427 15 US-10-369-493-11890 Sequence 11890, A  
715 19 100.0 427 15 US-10-104-047-2262 Sequence 2262, Ap  
716 19 100.0 428 12 US-10-403-571-110 Sequence 110, Appl  
717 19 100.0 430 15 US-10-369-493-17335 Sequence 17335, A  
718 19 100.0 431 12 US-10-282-132A-47233 Sequence 47233, A  
719 19 100.0 431 12 US-10-282-132A-51920 Sequence 51920, A  
720 19 100.0 431 12 US-10-424-599-212351 Sequence 212351, A  
721 19 100.0 431 12 US-10-041-018-260 Sequence 260, Appl  
722 19 100.0 431 15 US-10-369-493-21530 Sequence 21530, A  
723 19 100.0 432 15 US-10-425-114-72646 Sequence 72646, A  
724 19 100.0 432 15 US-10-369-493-8573 Sequence 8573, Ap  
725 19 100.0 433 14 US-10-128-714-8545 Sequence 8545, Ap  
726 19 100.0 433 15 US-10-369-493-10203 Sequence 10203, A  
727 19 100.0 434 12 US-10-282-132A-66904 Sequence 66904, A  
728 19 100.0 435 12 US-10-627-476-238 Sequence 238, Appl  
729 19 100.0 436 12 US-10-282-132A-50291 Sequence 50291, A  
730 19 100.0 436 12 US-10-425-114-46821 Sequence 46821, A  
731 19 100.0 436 12 US-10-041-018-235 Sequence 235, Appl  
732 19 100.0 436 15 US-10-369-493-21378 Sequence 21378, A  
733 19 100.0 437 9 US-09-519-532-2 Sequence 2, Appl  
734 19 100.0 437 9 US-09-738-626-4226 Sequence 4226, Ap  
735 19 100.0 437 12 US-09-746-660A-2 Sequence 2, Appl  
736 19 100.0 437 12 US-10-282-132A-42940 Sequence 42940, A  
737 19 100.0 437 12 US-10-282-132A-73421 Sequence 73421, A  
738 19 100.0 437 12 US-10-450-055-6 Sequence 6, Appl  
739 19 100.0 437 14 US-10-177-293-457 Sequence 457, Appl  
740 19 100.0 438 12 US-10-295-027-1180 Sequence 1180, Ap  
741 19 100.0 438 12 US-10-282-132A-56153 Sequence 56153, A  
742 19 100.0 438 12 US-10-282-132A-56161 Sequence 56161, A  
743 19 100.0 438 12 US-10-282-132A-75202 Sequence 75202, A  
744 19 100.0 438 12 US-10-282-132A-75609 Sequence 75609, A  
745 19 100.0 439 11 US-09-764-875-930 Sequence 930, Appl

Sequence 48732, A  
Sequence 1140, Ap  
Sequence 69, Appl  
Sequence 285027, A  
Sequence 69, Appl  
Sequence 440, Appl  
Sequence 65349, A  
Sequence 65787, A  
Sequence 68742, A  
Sequence 12502, A  
Sequence 4737, Ap  
Sequence 4882, Ap  
Sequence 7496, Ap  
Sequence 274539, A  
Sequence 56059, A  
Sequence 3345, Ap  
Sequence 3345, Ap  
Sequence 29, Appl  
Sequence 20475, A  
Sequence 13961, A  
Sequence 5303, Ap  
Sequence 13388, A  
Sequence 12820, A  
Sequence 71223, A  
Sequence 56431, A  
Sequence 8, Appl  
Sequence 22, Appl  
Sequence 5557, Ap  
Sequence 45014, A  
Sequence 214840, A  
Sequence 14870, A  
Sequence 7776, Ap  
Sequence 20, Appl  
Sequence 48506, A  
Sequence 2977, Ap  
Sequence 11890, A  
Sequence 2262, Ap  
Sequence 110, Appl  
Sequence 17335, A  
Sequence 47233, A  
Sequence 51920, A  
Sequence 212351, A  
Sequence 260, Appl  
Sequence 21530, A  
Sequence 72646, A  
Sequence 8573, Ap  
Sequence 8545, Ap  
Sequence 10203, A  
Sequence 66904, A  
Sequence 238, Appl  
Sequence 50291, A  
Sequence 46821, A  
Sequence 235, Appl  
Sequence 21378, A  
Sequence 2, Appl  
Sequence 4226, Ap  
Sequence 2, Appl  
Sequence 42940, A  
Sequence 73421, A  
Sequence 6, Appl  
Sequence 457, Appl  
Sequence 1180, Ap  
Sequence 56153, A  
Sequence 56161, A  
Sequence 75202, A  
Sequence 75609, A  
Sequence 930, Appl

Sequence 55457, A  
Sequence 77845, A  
Sequence 57263, A  
Sequence 161614, A  
Sequence 161615, A  
Sequence 226388, A  
Sequence 56949, A  
Sequence 124, Appl  
Sequence 73, Appl  
Sequence 45092, A  
Sequence 74539, A  
Sequence 17, Appl  
Sequence 48392, A  
Sequence 52177, A  
Sequence 4, Appl  
Sequence 72069, A  
Sequence 72107, A  
Sequence 17, Appl  
Sequence 12559, A  
Sequence 48954, A  
Sequence 266964, A  
Sequence 1420, Ap  
Sequence 39657, A  
Sequence 2490, Ap  
Sequence 5635, Ap  
Sequence 72044, A  
Sequence 74651, A  
Sequence 63654, A  
Sequence 72, Appl  
Sequence 55920, A  
Sequence 78329, A  
Sequence 3204, Ap  
Sequence 110, Appl  
Sequence 71, Appl  
Sequence 67190, A  
Sequence 11073, A  
Sequence 13584, A  
Sequence 8104, Ap  
Sequence 45167, A  
Sequence 56969, A  
Sequence 8005, Ap  
Sequence 2540, Ap  
Sequence 50440, A  
Sequence 186, Ap  
Sequence 366524, A  
Sequence 19, Appl  
Sequence 63275, A  
Sequence 61, Appl  
Sequence 1416, Ap  
Sequence 65, Appl  
Sequence 77, Appl  
Sequence 7098, Ap  
Sequence 76782, A  
Sequence 72794, A  
Sequence 27, Appl  
Sequence 251597, A  
Sequence 67114, A  
Sequence 68, Appl  
Sequence 62083, A  
Sequence 40744, A  
Sequence 39831, A  
Sequence 237731, A  
Sequence 14307, A  
Sequence 38072, A  
Sequence 108, Appl  
Sequence 166523, A  
Sequence 63937, A  
Sequence 108, Appl  
Sequence 166520, A  
Sequence 48208, A  
Sequence 42223, A

819	19	100.0	485	9	US-09-769-864-1	Sequence 1, Appli	892	19	100.0	518	12	US-10-424-599-172544	Sequence 172544,
820	19	100.0	485	9	US-09-769-864-2	Sequence 2, Appli	893	19	100.0	522	15	US-10-099-322-124	Sequence 124, App
821	19	100.0	485	9	US-09-769-864-6	Sequence 6, Appli	894	19	100.0	522	15	US-10-099-322-125	Sequence 125, App
822	19	100.0	485	9	US-09-769-864-7	Sequence 7, Appli	895	19	100.0	522	15	US-10-044-564-124	Sequence 124, App
823	19	100.0	485	9	US-09-769-864-8	Sequence 8, Appli	896	19	100.0	522	15	US-10-044-564-125	Sequence 125, App
824	19	100.0	485	9	US-09-854-346-2	Sequence 2, Appli	897	19	100.0	523	15	US-10-099-322-122	Sequence 122, App
825	19	100.0	485	9	US-09-854-346-4	Sequence 4, Appli	898	19	100.0	523	15	US-10-099-322-123	Sequence 123, App
826	19	100.0	485	9	US-09-854-346-12	Sequence 12, Appli	899	19	100.0	523	15	US-10-044-564-122	Sequence 122, App
827	19	100.0	485	9	US-09-854-346-13	Sequence 13, Appli	900	19	100.0	523	15	US-10-044-564-123	Sequence 123, App
828	19	100.0	485	9	US-09-854-346-13	Sequence 13, Appli	901	19	100.0	524	12	US-10-425-114-62567	Sequence 62567, A
829	19	100.0	485	9	US-09-302-188A-1	Sequence 1, Appli	902	19	100.0	524	12	US-10-425-114-68132	Sequence 68132, A
830	19	100.0	485	9	US-09-302-188A-7	Sequence 7, Appli	903	19	100.0	524	12	US-10-425-114-68132	Sequence 68132, A
831	19	100.0	485	9	US-09-302-188A-7	Sequence 7, Appli	904	19	100.0	524	12	US-10-403-161-34	Sequence 34, Appli
832	19	100.0	485	9	US-09-318-543-2	Sequence 2, Appli	905	19	100.0	524	12	US-10-403-161-40	Sequence 40, Appli
833	19	100.0	485	9	US-09-318-543-4	Sequence 4, Appli	906	19	100.0	524	12	US-10-403-161-42	Sequence 42, Appli
834	19	100.0	485	9	US-09-318-543-12	Sequence 12, Appli	907	19	100.0	524	15	US-10-099-322-34	Sequence 34, Appli
835	19	100.0	485	9	US-09-795-211-1	Sequence 1, Appli	908	19	100.0	524	15	US-10-099-322-121	Sequence 121, App
836	19	100.0	485	9	US-09-795-211-2	Sequence 2, Appli	909	19	100.0	524	15	US-10-044-564-34	Sequence 34, Appli
837	19	100.0	485	10	US-09-925-576C-2	Sequence 2, Appli	910	19	100.0	526	9	US-09-815-242-13696	Sequence 13696, A
838	19	100.0	485	10	US-09-925-576C-4	Sequence 4, Appli	911	19	100.0	526	12	US-10-282-122A-73677	Sequence 73677, A
839	19	100.0	485	10	US-09-925-576C-12	Sequence 12, Appli	912	19	100.0	526	12	US-10-425-114-46419	Sequence 46419, A
840	19	100.0	485	10	US-09-925-576C-13	Sequence 13, Appli	913	19	100.0	527	12	US-10-425-114-43548	Sequence 43548, A
841	19	100.0	485	12	US-10-665-667-1	Sequence 1, Appli	914	19	100.0	527	12	US-10-425-114-43548	Sequence 43548, A
842	19	100.0	485	12	US-10-665-667-2	Sequence 2, Appli	915	19	100.0	529	12	US-10-403-161-36	Sequence 36, Appli
843	19	100.0	485	12	US-10-665-667-6	Sequence 6, Appli	916	19	100.0	530	15	US-10-403-161-38	Sequence 38, Appli
844	19	100.0	485	12	US-10-665-667-8	Sequence 8, Appli	917	19	100.0	531	12	US-10-369-493-19139	Sequence 19139, A
845	19	100.0	485	12	US-10-025-648-2	Sequence 2, Appli	918	19	100.0	531	12	US-10-425-114-45814	Sequence 45814, A
846	19	100.0	485	12	US-10-025-648-7	Sequence 7, Appli	919	19	100.0	533	12	US-10-282-122A-61253	Sequence 61253, A
847	19	100.0	485	12	US-10-327-837-24	Sequence 24, Appli	920	19	100.0	533	14	US-10-156-761-9651	Sequence 9651, Ap
848	19	100.0	485	12	US-10-327-837-25	Sequence 25, Appli	921	19	100.0	535	12	US-10-282-122A-60086	Sequence 60086, A
849	19	100.0	485	12	US-10-327-837-1	Sequence 1, Appli	922	19	100.0	535	12	US-10-282-122A-60086	Sequence 60086, A
850	19	100.0	485	12	US-10-327-837-6	Sequence 6, Appli	923	19	100.0	536	12	US-10-425-114-62633	Sequence 62633, A
851	19	100.0	485	12	US-10-327-837-7	Sequence 7, Appli	924	19	100.0	536	15	US-10-425-114-69007	Sequence 69007, A
852	19	100.0	485	12	US-10-327-837-8	Sequence 8, Appli	925	19	100.0	537	12	US-10-369-493-7664	Sequence 7664, Ap
853	19	100.0	485	12	US-10-327-837-24	Sequence 24, Appli	926	19	100.0	537	12	US-10-282-122A-75618	Sequence 75618, A
854	19	100.0	485	12	US-10-327-837-25	Sequence 25, Appli	927	19	100.0	537	12	US-10-425-114-64000	Sequence 64000, A
855	19	100.0	485	14	US-10-208-812-5	Sequence 5, Appli	928	19	100.0	537	15	US-10-120-801-12	Sequence 12, Appli
856	19	100.0	485	14	US-10-184-771-12	Sequence 12, Appli	929	19	100.0	538	12	US-10-282-122A-4952	Sequence 4952, A
857	19	100.0	485	15	US-10-369-493-4906	Sequence 4906, Ap	930	19	100.0	538	12	US-10-425-114-70508	Sequence 70508, A
858	19	100.0	487	12	US-10-403-161-44	Sequence 44, Appli	931	19	100.0	539	12	US-10-282-122A-57265	Sequence 57265, A
859	19	100.0	487	15	US-10-369-493-18040	Sequence 18040, A	932	19	100.0	539	12	US-10-425-114-46493	Sequence 46493, A
860	19	100.0	487	12	US-10-425-114-45580	Sequence 45580, A	933	19	100.0	539	14	US-10-156-761-11731	Sequence 11731, A
861	19	100.0	481	10	US-09-863-776-34	Sequence 34, Appli	934	19	100.0	540	9	US-10-230-026-34	Sequence 34, Appli
862	19	100.0	491	12	US-10-424-599-149190	Sequence 149190,	935	19	100.0	540	15	US-09-738-626-4450	Sequence 4450, Ap
863	19	100.0	492	12	US-09-881-752A-142	Sequence 142, App	936	19	100.0	540	15	US-10-369-493-3921	Sequence 3921, Ap
864	19	100.0	492	12	US-10-283-132A-58846	Sequence 58846, A	937	19	100.0	543	12	US-10-282-122A-77151	Sequence 77151, A
865	19	100.0	493	12	US-09-918-951-5	Sequence 5, Appli	938	19	100.0	543	15	US-10-369-493-10792	Sequence 10792, A
866	19	100.0	493	12	US-10-614-076-70	Sequence 70, Appli	939	19	100.0	546	12	US-10-087-192-843	Sequence 843, App
867	19	100.0	493	12	US-10-369-493-20452	Sequence 20452, A	940	19	100.0	547	12	US-10-282-122A-53681	Sequence 53681, A
868	19	100.0	494	15	US-10-369-493-5803	Sequence 5803, Ap	941	19	100.0	547	14	US-10-043-344-149	Sequence 149, App
869	19	100.0	496	12	US-10-369-493-5803	Sequence 5803, Ap	942	19	100.0	548	15	US-10-282-122A-65960	Sequence 65960, A
870	19	100.0	497	15	US-10-369-493-11919	Sequence 11919, A	943	19	100.0	548	15	US-10-369-493-6495	Sequence 6495, Ap
871	19	100.0	499	9	US-10-369-493-11919	Sequence 11919, A	944	19	100.0	549	12	US-10-369-493-6496	Sequence 6496, Ap
872	19	100.0	499	12	US-10-283-132A-58172	Sequence 58172, A	945	19	100.0	549	15	US-09-815-242-11206	Sequence 11206, A
873	19	100.0	499	12	US-10-424-599-149746	Sequence 149746, A	946	19	100.0	549	12	US-10-282-122A-58451	Sequence 58451, A
874	19	100.0	499	15	US-10-369-493-8665	Sequence 8665, Ap	947	19	100.0	549	15	US-10-282-122A-66993	Sequence 66993, A
875	19	100.0	501	12	US-10-282-122A-63359	Sequence 63359, A	948	19	100.0	550	9	US-10-369-493-5478	Sequence 5478, Ap
876	19	100.0	503	12	US-10-424-599-162427	Sequence 162427, A	949	19	100.0	550	11	US-09-815-242-10499	Sequence 10499, A
877	19	100.0	504	12	US-10-424-599-162427	Sequence 162427, A	950	19	100.0	550	12	US-09-833-245-1284	Sequence 1284, Ap
878	19	100.0	504	12	US-10-424-599-162427	Sequence 162427, A	951	19	100.0	550	12	US-10-282-122A-77028	Sequence 77028, A
879	19	100.0	504	12	US-10-424-599-162427	Sequence 162427, A	952	19	100.0	550	14	US-10-336-049-2	Sequence 4988, A
880	19	100.0	504	12	US-10-424-599-162427	Sequence 162427, A	953	19	100.0	550	14	US-10-091-342-2	Sequence 2, Appli
881	19	100.0	504	12	US-10-424-599-162427	Sequence 162427, A	954	19	100.0	550	12	US-10-091-342-2	Sequence 2, Appli
882	19	100.0	505	15	US-10-369-493-8682	Sequence 8682, Ap	955	19	100.0	552	12	US-10-282-122A-59108	Sequence 59108, A
883	19	100.0	506	9	US-10-403-671-9	Sequence 9, Appli	956	19	100.0	553	12	US-10-282-122A-70501	Sequence 70501, A
884	19	100.0	506	15	US-10-403-671-9	Sequence 9, Appli	957	19	100.0	553	15	US-10-282-122A-71716	Sequence 71716, A
885	19	100.0	506	15	US-10-403-671-9	Sequence 9, Appli	958	19	100.0	553	15	US-10-313-371A-2	Sequence 2, Appli
886	19	100.0	506	15	US-10-403-671-9	Sequence 9, Appli	959	19	100.0	556	15	US-10-369-493-5442	Sequence 5442, Ap
887	19	100.0	515	12	US-10-424-599-146129	Sequence 146129, A	960	19	100.0	556	15	US-10-369-493-11974	Sequence 11974, A
888	19	100.0	515	12	US-10-424-599-146129	Sequence 146129, A	961	19	100.0	558	12	US-10-363-616-374	Sequence 374, App
889	19	100.0	516	9	US-09-865-676A-2	Sequence 2, Appli	962	19	100.0	558	15	US-10-369-493-6392	Sequence 6392, Ap
890	19	100.0	516	9	US-09-865-676A-2	Sequence 2, Appli	963	19	100.0	559	9	US-09-925-300-1479	Sequence 1479, Ap
891	19	100.0	516	16	US-10-399-161-8	Sequence 8, Appli	964	19	100.0	559	12	US-10-282-122A-65056	Sequence 65056, A

965 19 100.0 559 12 US-10-282-122A-75373  
 966 19 100.0 559 12 US-10-282-122A-76352  
 967 19 100.0 561 12 US-10-282-122A-65931  
 968 19 100.0 561 15 US-10-369-493-21529  
 969 19 100.0 564 16 US-10-389-566-19659  
 970 19 100.0 565 12 US-10-380-565-34  
 971 19 100.0 566 14 US-10-100-957A-142  
 972 19 100.0 566 14 US-10-211-088-19  
 973 19 100.0 567 12 US-10-425-114-45750  
 974 19 100.0 568 12 US-10-282-122A-65302  
 975 19 100.0 573 9 US-09-764-868-671  
 976 19 100.0 573 11 US-09-784-875-1205  
 977 19 100.0 573 12 US-10-425-114-40721  
 978 19 100.0 574 12 US-10-425-114-42248  
 979 19 100.0 575 12 US-10-424-599-186118  
 980 19 100.0 575 12 US-10-312-352-8  
 981 19 100.0 575 15 US-10-120-801-60  
 982 19 100.0 576 12 US-10-335-977-7018  
 983 19 100.0 577 9 US-09-815-242-11305  
 984 19 100.0 577 12 US-10-282-122A-58657  
 985 19 100.0 577 12 US-10-335-977-7019  
 986 19 100.0 580 9 US-09-841-132-598  
 987 19 100.0 580 12 US-10-282-122A-54815  
 988 19 100.0 580 15 US-10-289-762-332  
 989 19 100.0 581 12 US-10-424-599-255268  
 990 19 100.0 584 15 US-10-108-260A-4380  
 991 19 100.0 585 12 US-10-431-449-2  
 992 19 100.0 585 12 US-10-431-449-4  
 993 19 100.0 585 12 US-10-431-449-6  
 994 19 100.0 587 12 US-10-424-599-198064  
 995 19 100.0 587 13 US-10-042-523-3  
 996 19 100.0 588 15 US-10-309-560-10  
 997 19 100.0 590 12 US-10-282-122A-65911  
 998 19 100.0 592 12 US-10-424-599-214104  
 999 19 100.0 592 12 US-10-335-977-5869  
 1000 19 100.0 595 12 US-10-282-122A-44744

## ALIGNMENTS

RESULT 1  
 S-09-791-378-343  
 Sequence 343, Application US/09791378  
 Patent No. US20020142303A1  
 GENERAL INFORMATION:  
 APPLICANT: Parekh, Rajesh  
 TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF  
 SCHIZOPHRENIA  
 FILE REFERENCE: 9195-061-999  
 CURRENT APPLICATION NUMBER: US/09/791,378  
 CURRENT FILING DATE: 2001-02-23  
 PRIOR APPLICATION NUMBER: 09/750,395  
 PRIOR FILING DATE: 2000-12-28  
 NUMBER OF SEQ ID NOS: 677  
 SOFTWARE: Patentin version 3.0  
 SEQ ID NO 343  
 LENGTH: 8  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 S-09-791-378-343

Query Match 100.0%; Score 19; DB 9; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
 ||||  
 b 2 VAEF 5

RESULT 2  
 S-09-791-393-206

; Sequence 206, Application US/09791393  
 ; Publication No. US20030032200A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Herath, Herath Mudiyanseelage Athula Chandrasiri  
 ; APPLICANT: Parekh, Rajesh Bhikhu  
 ; APPLICANT: Rohlf, Christian  
 ; TITLE OF INVENTION: Proteins, Genes and Their Use for  
 ; TITLE OF INVENTION: Diagnosis and Treatment of Bipolar Affective Disorder (BAD)  
 ; TITLE OF INVENTION: and Unipolar Depression  
 ; FILE REFERENCE: 2543-1-001 N1  
 ; CURRENT APPLICATION NUMBER: US/09/791,393  
 ; CURRENT FILING DATE: 2002-01-02  
 ; EARLIER APPLICATION NUMBER: GB 0004412.3  
 ; EARLIER FILING DATE: 2000-02-24  
 ; EARLIER APPLICATION NUMBER: GB 0030050.9  
 ; EARLIER FILING DATE: 2000-12-08  
 ; EARLIER APPLICATION NUMBER: US 60/254,830  
 ; EARLIER FILING DATE: 2000-12-12  
 ; NUMBER OF SEQ ID NOS: 308  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 206  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: homo sapien  
 US-09-791-393-206

Query Match 100.0%; Score 19; DB 10; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 Db 2 VAEF 5

## RESULT 3

US-09-791-389-206  
 ; Sequence 206, Application US/09791389  
 ; Publication No. US20030032773A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Herath, Herath Mudiyanseelage Athula Chandrasiri  
 ; APPLICANT: Parekh, Rajesh Bhikhu  
 ; APPLICANT: Rohlf, Christian  
 ; APPLICANT: Terrett, Jonathan Alexander  
 ; APPLICANT: Tyson, Kerry Louise  
 ; TITLE OF INVENTION: Proteins, Genes and Their Use for  
 ; TITLE OF INVENTION: Diagnosis and Treatment of Bipolar Affective Disorder (BAD)  
 ; TITLE OF INVENTION: and Unipolar Depression  
 ; FILE REFERENCE: 2543-1-001 N2  
 ; CURRENT APPLICATION NUMBER: US/09/791,389  
 ; CURRENT FILING DATE: 2001-02-23  
 ; PRIOR APPLICATION NUMBER: GB 0004412.3  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: GB 0030050.9  
 ; PRIOR FILING DATE: 2000-12-08  
 ; PRIOR APPLICATION NUMBER: US 60/254,830  
 ; PRIOR FILING DATE: 2000-12-12  
 ; NUMBER OF SEQ ID NOS: 308  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 206  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: homo sapien  
 US-09-791-389-206

Query Match 100.0%; Score 19; DB 10; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
 ||||  
 Db 2 VAEF 5

SULT 4  
 -10-128-711-86  
 Sequence 86, Application US/10128711  
 Publication No. US20030095634A1  
 GENERAL INFORMATION:  
 APPLICANT: VITIELLO, Maria A.  
 CHESTNUT, Robert W.  
 SETTE, Alessandro D.  
 CELIS, Esteban  
 GRAY, Howard  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING CTL IMMUNITY  
 NUMBER OF SEQUENCES: 153  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend Kourie and Crew  
 STREET: Steuart Street Tower, One Market Plaza  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: US  
 ZIP: 94105-1493  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/128,711  
 FILING DATE: 22-Apr-2002  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/197,484  
 FILING DATE: 16-FEB-1994  
 APPLICATION NUMBER: US 07/935,811  
 FILING DATE: 26-AUG-1992  
 APPLICATION NUMBER: US 07/874,491  
 FILING DATE: 27-APR-1992  
 APPLICATION NUMBER: US 07/827,682  
 FILING DATE: 29-JAN-1992  
 APPLICATION NUMBER: US 07/749,568  
 FILING DATE: 26-AUG-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Parmelee, Steven W.  
 REGISTRATION NUMBER: 31,990  
 REFERENCE/DOCKET NUMBER: 14137-26-4  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 467-9600  
 TELEFAX: (206) 623-6793  
 INFORMATION FOR SEQ ID NO: 86:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: unknown  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 SEQUENCE DESCRIPTION: SEQ ID NO: 86:  
 -10-128-711-86  
 Query Match 100.0%; Score 19; DB 14; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 VAEF 4  
 2 VAEF 5  
 RESULT 5  
 -10-128-711-145  
 Sequence 145, Application US/10128711  
 Publication No. US20030095634A1  
 GENERAL INFORMATION:  
 APPLICANT: VITIELLO, Maria A.

CHESTNUT, Robert W.  
 SETTE, Alessandro D.  
 CELIS, Esteban  
 GRAY, Howard  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING CTL IMMUNITY  
 NUMBER OF SEQUENCES: 153  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend Kourie and Crew  
 STREET: Steuart Street Tower, One Market Plaza  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: US  
 ZIP: 94105-1493  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/128,711  
 FILING DATE: 22-Apr-2002  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/197,484  
 FILING DATE: 16-FEB-1994  
 APPLICATION NUMBER: US 07/935,811  
 FILING DATE: 26-AUG-1992  
 APPLICATION NUMBER: US 07/874,491  
 FILING DATE: 27-APR-1992  
 APPLICATION NUMBER: US 07/827,682  
 FILING DATE: 29-JAN-1992  
 APPLICATION NUMBER: US 07/749,568  
 FILING DATE: 26-AUG-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Parmelee, Steven W.  
 REGISTRATION NUMBER: 31,990  
 REFERENCE/DOCKET NUMBER: 14137-26-4  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 467-9600  
 TELEFAX: (206) 623-6793  
 INFORMATION FOR SEQ ID NO: 145:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: unknown  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 SEQUENCE DESCRIPTION: SEQ ID NO: 145:  
 US-10-128-711-145  
 Query Match 100.0%; Score 19; DB 14; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 VAEF 4  
 2 VAEF 5  
 RESULT 6  
 US-09-908-943A-197  
 Sequence 197, Application US/0908943A  
 Publication No. US20030017991A1  
 GENERAL INFORMATION:  
 APPLICANT: Yan, Riqiang  
 APPLICANT: Tomaseilli, Alfredo G.  
 APPLICANT: Gurney, Mark E.  
 APPLICANT: Emmons, Thomas L.  
 APPLICANT: Bienkowski, Mike J.  
 APPLICANT: Heinrichson, Robert L.  
 TITLE OF INVENTION: SUBSTRATES AND ASSAYS FOR BETA-SECRETASE ACTIVITY  
 FILE REFERENCE: 29915/00281A.US1

CURRENT APPLICATION NUMBER: US/09/908,943A

CURRENT FILING DATE: 2001-07-19

PRIOR APPLICATION NUMBER: 60/219,795

PRIOR FILING DATE: 2000-07-19

NUMBER OF SEQ ID NOS: 197

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 197

LENGTH: 10

TYPE: PRT

ORGANISM: synthetic peptide sequence

FEATURE:

NAME/KEY: SITE

LOCATION: (4)..(4)

OTHER INFORMATION: amino acid at position 4 has been derivatized with a statine

NAME/KEY: SITE

LOCATION: (10)..(10)

OTHER INFORMATION: amino acid at position 10 has been derivatized with Bodipy FL

US-09-908-943A-197

Query Match 100.0%; Score 19; DB 12; Length 10;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4

Db 5 VAEF 8

RESULT 7

US-09-908-943A-196

Sequence 196, Application US/09908943A

Publication No. US20030017991A1

GENERAL INFORMATION:

APPLICANT: Yan, Rigdang

APPLICANT: Tomasselli, Alfredo G.

APPLICANT: Gurney, Mark E.

APPLICANT: Emmons, Thomas L.

APPLICANT: Bienkowski, Mike J.

APPLICANT: Heinrikson, Robert L.

TITLE OF INVENTION: SUBSTRATES AND ASSAYS FOR BETA-SECRETASE ACTIVITY

FILE REFERENCE: 29915/00281A.US1

CURRENT APPLICATION NUMBER: US/09/908,943A

CURRENT FILING DATE: 2001-07-19

PRIOR APPLICATION NUMBER: 60/219,795

PRIOR FILING DATE: 2000-07-19

NUMBER OF SEQ ID NOS: 197

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 196

LENGTH: 12

TYPE: PRT

ORGANISM: Artificial sequence

FEATURE:

OTHER INFORMATION: Description of artificial sequence: synthetic peptide sequence

NAME/KEY: SITE

LOCATION: (4)..(4)

OTHER INFORMATION: amino acid at position 4 has been derivatized with a statine

US-09-908-943A-196

Query Match

Best Local Similarity 100.0%; Score 19; DB 12; Length 12;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4

Db 5 VAEF 8

RESULT 8

US-10-032-818-34

Sequence 34, Application US/10032818

Publication No. US20030092629A1

GENERAL INFORMATION:

APPLICANT: Tang, Jordan J.N.

APPLICANT: Koelsch, Gerald

APPLICANT: Ghost, Arun K.

TITLE OF INVENTION: Inhibitors of Memapsin 2 and Use Thereof

FILE REFERENCE: 2932.1006-007

CURRENT APPLICATION NUMBER: US/10/032,818

CURRENT FILING DATE: 2001-12-28

PRIOR APPLICATION NUMBER: US 60/375,756

PRIOR FILING DATE: 2001-03-14

PRIOR APPLICATION NUMBER: US 60/258,705

PRIOR FILING DATE: 2000-12-28

NUMBER OF SEQ ID NOS: 83

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 34

LENGTH: 12

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Synthetic peptide

US-10-032-818-34

Query Match

Best Local Similarity 100.0%; Score 19; DB 14; Length 12;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4

Db 5 VAEF 8

RESULT 9

US-10-239-313A-379

Sequence 379, Application US/10239313A

Publication No. US20030175285A1

GENERAL INFORMATION:

APPLICANT: KLINGUER - HAMOUR, Christine

APPLICANT: CORVAIA, Nathalie

APPLICANT: BECK, Alain

APPLICANT: GOETSCH, Liliane

TITLE OF INVENTION: MOLECULE OF PHARMACEUTICAL INTEREST COMPRISING AT ITS

TITLE OF INVENTION: N-TERMINAL A GLUTAMIC ACID OR A GLUTAMINE IN THE FORM

FILE REFERENCE: 343 727 US

CURRENT APPLICATION NUMBER: US/10/239,313A

CURRENT FILING DATE: 2002-09-19

PRIOR APPLICATION NUMBER: FR 00/03711

PRIOR FILING DATE: 2000-03-23

PRIOR APPLICATION NUMBER: PCT 01/70772

PRIOR FILING DATE: 2001-03-22

NUMBER OF SEQ ID NOS: 697

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 379

LENGTH: 13

TYPE: PRT

ORGANISM: Mycobacterium leprae

US-10-239-313A-379

Query Match

Best Local Similarity 100.0%; Score 19; DB 14; Length 13;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4

Db 4 VAEF 7

RESULT 10

US-10-119-528-7

Sequence 7, Application US/10119528

Publication No. US20030175722A1

GENERAL INFORMATION:

APPLICANT: Mann, M.

APPLICANT: Mortensen, P.

TITLE OF INVENTION: METHODS AND SYSTEMS FOR SEARCHING GENOMIC DATABASES

FILE REFERENCE: MDSP-P01-004  
CURRENT APPLICATION NUMBER: US/10/119,528  
CURRENT FILING DATE: 2002-04-09  
PRIOR APPLICATION NUMBER: 60/282,551  
PRIOR FILING DATE: 2001-04-09  
PRIOR APPLICATION NUMBER: 60/285,362  
PRIOR FILING DATE: 2001-04-20  
NUMBER OF SEQ ID NOS: 112  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 7  
LENGTH: 15  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
-10-119-528-7

Query Match 100.0%; Score 19; DB 14; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2e+02; 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|  
|  
|  
|  
DB 10 VAEF 13

SULT 11  
-09-171-432A-62  
Sequence 62, Application US/09171432A  
Publication No. US20030187184A1  
GENERAL INFORMATION:  
APPLICANT: Khudiyakov, Yuri E.  
TITLE OF INVENTION: Antigenically Reactive Regions of the  
Hepatitis A Virus Polyprotein  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick Stockton LLP  
STREET: 3424 Peachtree Road, N.E.  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30326

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/171,432A  
FILING DATE: 23-NOV-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Greene, Jamie L.  
REGISTRATION NUMBER: 32,467  
REFERENCE/DOCKET NUMBER: 03063-0231US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 949-2400  
TELEFAX: (404) 949-2499  
INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..20  
OTHER INFORMATION: /label= YK-1368

US-09-171-432A-62  
US-09-171-432A-65  
Sequence 65, Application US/09171432A  
Publication No. US20030187184A1  
GENERAL INFORMATION:  
APPLICANT: Fields, Howard A.

Query Match 100.0%; Score 19; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02; 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|  
|  
|  
|  
DB 4 VAEF 7

RESULT 13  
US-09-171-432A-65  
Sequence 65, Application US/09171432A  
Publication No. US20030187184A1  
GENERAL INFORMATION:  
APPLICANT: Fields, Howard A.

Query Match 100.0%; Score 19; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02; 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|  
|  
|  
|  
DB 13 VAEF 16

RESULT 12  
US-09-171-432A-63  
Sequence 63, Application US/09171432A  
Publication No. US20030187184A1  
GENERAL INFORMATION:  
APPLICANT: Khudiyakov, Yuri E.  
TITLE OF INVENTION: Antigenically Reactive Regions of the  
Hepatitis A Virus Polyprotein  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick Stockton LLP  
STREET: 3424 Peachtree Road, N.E.  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30326

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/171,432A  
FILING DATE: 23-NOV-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Greene, Jamie L.  
REGISTRATION NUMBER: 32,467  
REFERENCE/DOCKET NUMBER: 03063-0231US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 949-2400  
TELEFAX: (404) 949-2499  
INFORMATION FOR SEQ ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..20  
OTHER INFORMATION: /label= YK-1369

Query Match 100.0%; Score 19; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02; 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

APPLICANT: Khudiyakov, Yuri E.  
TITLE OF INVENTION: Antigenically Reactive Regions of the  
Hepatitis A Virus Polyprotein  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick Stockton LLP  
STREET: 3424 Peachtree Road, N.E.  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30326

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/171,432A  
FILING DATE: 23-NOV-1998

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Greene, Jamie L.  
REGISTRATION NUMBER: 32,467  
REFERENCE/DOCKET NUMBER: 03063-0231US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 949-2400  
TELEFAX: (404) 949-2499  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal

FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..25  
OTHER INFORMATION: /label=YK-1832

S-09-171-432A-65

Query Match 100.0%; Score 19; DB 10; Length 25;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4

b 3 VAEF 6

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 47201  
LENGTH: 28  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC005772.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.46  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.67  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.61  
US-09-864-761-47201

Query Match 100.0%; Score 19; DB 9; Length 28;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAEF 4

Db 4 VAEF 7

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

RESULT 15

US-10-424-599-216297  
Sequence 216297, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 216297  
LENGTH: 28  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_37345C.1.pdp  
US-10-424-599-216297

Query Match 100.0%; Score 19; DB 12; Length 28;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 VAEF 4  
14 VAEF 13

SULT 16  
-10-043-344-69  
Sequence 69, Application US/10043344  
Publication No. US20030088086A1  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Harkness, Robin E.  
APPLICANT: Schryvers, Anthony B.  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Murdin, Andrew D.  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES  
FILE REFERENCE: 1038-1221 MIS  
CURRENT APPLICATION NUMBER: US/10/043,344  
CURRENT FILING DATE: 2002-07-01  
PRIOR APPLICATION NUMBER: 08/649,518  
PRIOR FILING DATE: 1996-05-17  
NUMBER OF SEQ ID NOS: 160  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 69  
LENGTH: 35  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
-10-043-344-69

Query Match 100.0%; Score 19; DB 14; Length 35;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 VAEF 4  
14 VAEF 17

SULT 17  
-10-242-355-570  
Sequence 570, Application US/10242355  
Publication No. US20030235831A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC003C1  
CURRENT APPLICATION NUMBER: US/10/242,355  
CURRENT FILING DATE: 2002-09-13  
PRIOR APPLICATION NUMBER: 09/764,897  
PRIOR FILING DATE: 2001-01-17  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: 60/214,886  
PRIOR FILING DATE: 2000-06-26  
PRIOR APPLICATION NUMBER: 60/217,487  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,758  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,963  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/217,496  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,447  
PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1267  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 570  
; LENGTH: 37  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-242-355-570

Query Match 100.0%; Score 19; DB 15; Length 37;  
Best Local Similarity 100.0%; Pred. No. 4.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 VAEF 4  
14 VAEF 17

RESULT 18  
US-10-430-752A-49  
Sequence 49, Application US/10430752A  
Publication No. US20040005669A1  
GENERAL INFORMATION:  
APPLICANT: Stahl, Stefan  
APPLICANT: Jonasson, Per  
APPLICANT: Nygren, Per-Ake  
APPLICANT: Uhlen, Mathias  
TITLE OF INVENTION: RECOMBINANT EXPRESSION OF INSULIN  
TITLE OF INVENTION: C-PEPTIDE  
FILE REFERENCE: 11541-003001  
CURRENT APPLICATION NUMBER: US/10/430,752A  
CURRENT FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: 09/485,286  
PRIOR FILING DATE: 2000-02-07  
PRIOR APPLICATION NUMBER: PCT/GB98/02382  
PRIOR FILING DATE: 1998-08-07  
PRIOR APPLICATION NUMBER: GB 9716790.2  
PRIOR FILING DATE: 1997-08-07  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 49  
LENGTH: 37  
TYPE: PRT  
ORGANISM: Verasper moseri  
US-10-430-752A-49

Query Match 100.0%; Score 19; DB 15; Length 37;  
Best Local Similarity 100.0%; Pred. No. 4.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 VAEF 4  
24 VAEF 27

RESULT 19  
US-10-430-752A-45  
Sequence 45, Application US/10430752A  
Publication No. US20040005669A1  
GENERAL INFORMATION:  
APPLICANT: Stahl, Stefan  
APPLICANT: Jonasson, Per  
APPLICANT: Nygren, Per-Ake  
APPLICANT: Uhlen, Mathias  
TITLE OF INVENTION: RECOMBINANT EXPRESSION OF INSULIN  
TITLE OF INVENTION: C-PEPTIDE  
FILE REFERENCE: 11541-003001  
CURRENT APPLICATION NUMBER: US/10/430,752A  
CURRENT FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: 09/485,286  
PRIOR FILING DATE: 2000-02-07

PRIOR APPLICATION NUMBER: PCT/GB98/02382

PRIOR FILING DATE: 1998-08-07

PRIOR APPLICATION NUMBER: GB 9716790.2

PRIOR FILING DATE: 1997-08-07

NUMBER OF SEQ ID NOS: 63

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 45

LENGTH: 38

TYPE: PRT

ORGANISM: Lophius piscatorius

S-10-430-752A-45

Query Match 100.0%; Score 19; DB 15; Length 38;

Best Local Similarity 100.0%; Pred. No. 5.1e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 VAEF 4

b 25 VAEF 28

RESULT 20

S-10-372-003A-71

Sequence 71, Application US/10372003A

Publication No. US20030215846A1

GENERAL INFORMATION:

APPLICANT: Watt, Paul

APPLICANT: Thomas, Wayne

APPLICANT: Hopkins, Richard

TITLE OF INVENTION: Methods of constructing and screening

TITLE OF INVENTION: diverse expression libraries

FILE REFERENCE: FBIC40.001CPI

CURRENT APPLICATION NUMBER: US/10/372,003A

CURRENT FILING DATE: 2003-02-21

PRIOR APPLICATION NUMBER: US 09/568,229

PRIOR FILING DATE: 2000-05-05

PRIOR APPLICATION NUMBER: US 60/132,711

PRIOR FILING DATE: 1999-05-05

NUMBER OF SEQ ID NOS: 81

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 71

LENGTH: 41

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: synthetic clone BGFI3 with first open reading

OTHER INFORMATION: frame

S-10-372-003A-71

Query Match 100.0%; Score 19; DB 15; Length 41;

Best Local Similarity 100.0%; Pred. No. 5.5e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 VAEF 4

b 28 VAEF 31

RESULT 21

S-09-864-761-37202

Sequence 37202, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Acomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Anncmax Sequence Listing Engine vers. 1.1

SEQ ID NO 37202

LENGTH: 43

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AL035530.11

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.86

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1

OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5

OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.95

OTHER INFORMATION: SWISSPROT HIT: P33485, EVALUE 3.60e+00

OTHER INFORMATION: EST\_HUMAN HIT: A1765888.1, EVALUE 3.00e-18

US-09-864-761-37202

Query Match 100.0%; Score 19; DB 9; Length 43;

Best Local Similarity 100.0%; Pred. No. 5.7e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAEF 4

Db 40 VAEF 43

RESULT 22

US-10-424-599-183032

Sequence 183032, Application US/10424599

Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J

APPLICANT: Kovalic David K

APPLICANT: Zhou Yinhua

APPLICANT: Cao Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53223)B

```

: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/006666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006685
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006653
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006652
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
: SEQ ID NO 38166
: LENGTH: 48
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL049833.3
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.96
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
: OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
: OTHER INFORMATION: SWISSPROT HIT: P11298, EVALUE 7.90e+00
: US-09-864-761-38166
:
: Query Match 100.0%; Score 19; DB 9; Length 48;
: Best Local Similarity 100.0%; Pred. No. 6.4e+02;
: Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: QY 1 VAEF 4
: DB 3 VAEF 6
:
: RESULT 25
: US-10-424-599-205632
: Sequence 205632, Application US/10424599
: Publication No. US20040031072A1
: GENERAL INFORMATION:
: APPLICANT: La Rosa Thomas J
: APPLICANT: Kovalic David K
: APPLICANT: Zhou Yihua
: APPLICANT: Cao Yongwei
: TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-21(53223)B
: CURRENT APPLICATION NUMBER: US/10/424,599
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 285684
: SEQ ID NO 205632
: LENGTH: 52

```

```

TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_27713C.1.pep
S-10-424-599-205632

Query Match      100.0%; Score 19; DB 12; Length 52;
Best Local Similarity 100.0%; Pred. No. 6.9e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0;

Y 1 VAEF 4
b 26 VAEF 29

RESULT 26
S-10-424-599-192247
Sequence 192247, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 192247
LENGTH: 55
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_1561C.1.pep
S-10-424-599-192247

Query Match      100.0%; Score 19; DB 12; Length 55;
Best Local Similarity 100.0%; Pred. No. 7.3e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0;

Y 1 VAEF 4
b 50 VAEF 53

RESULT 27
S-10-424-599-233502
Sequence 233502, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 233502
LENGTH: 55
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_52879C.1.pep
S-10-424-599-233502

Query Match      100.0%; Score 19; DB 12; Length 55;
Best Local Similarity 100.0%; Pred. No. 7.3e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0;

Y 1 VAEF 4
b 24 VAEF 27

RESULT 28
US-10-372-003A-63
Sequence 63, Application US/10372003A
Publication No. US20030215846A1
GENERAL INFORMATION:
APPLICANT: Watt, Paul
APPLICANT: Thomas, Wayne
APPLICANT: Hopkins, Richard
TITLE OF INVENTION: Methods of constructing and screening
FILE REFERENCE: FBIC40.001CPI
CURRENT APPLICATION NUMBER: US/10/372,003A
CURRENT FILING DATE: 2003-02-21
PRIOR APPLICATION NUMBER: US 09/568,229
PRIOR FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: US 60/132,711
PRIOR FILING DATE: 1999-05-05
NUMBER OF SEQ ID NOS: 81
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 63
LENGTH: 55
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: synthetic clone BGF05 with first open reading
OTHER INFORMATION: frame
US-10-372-003A-63

Query Match      100.0%; Score 19; DB 15; Length 55;
Best Local Similarity 100.0%; Pred. No. 7.3e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0;

Y 1 VAEF 4
b 28 VAEF 31

RESULT 29
US-10-372-003A-67
Sequence 67, Application US/10372003A
Publication No. US20030215846A1
GENERAL INFORMATION:
APPLICANT: Watt, Paul
APPLICANT: Thomas, Wayne
APPLICANT: Hopkins, Richard
TITLE OF INVENTION: Methods of constructing and screening
FILE REFERENCE: FBIC40.001CPI
CURRENT APPLICATION NUMBER: US/10/372,003A
CURRENT FILING DATE: 2003-02-21
PRIOR APPLICATION NUMBER: US 09/568,229
PRIOR FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: US 60/132,711
PRIOR FILING DATE: 1999-05-05
NUMBER OF SEQ ID NOS: 81
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 67
LENGTH: 55
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: synthetic clone BGF06 with first open reading
OTHER INFORMATION: frame
US-10-372-003A-67

Query Match      100.0%; Score 19; DB 15; Length 55;
Best Local Similarity 100.0%; Pred. No. 7.3e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0;
```

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
|||||  
28 VAEF 31

SULT 30  
-10-219-329-4  
Sequence 4, Application US/10219329  
Publication No. US20030096757A1  
GENERAL INFORMATION:  
APPLICANT: Quirk, Stephen  
APPLICANT: Weart, Ilona f.  
TITLE OF INVENTION: Anti-Cancer and Wound Healing Compounds  
FILE REFERENCE: 1443.035W01  
CURRENT APPLICATION NUMBER: US/10/219,329  
CURRENT FILING DATE: 2002-08-15  
PRIOR APPLICATION NUMBER: US 10/032,376  
PRIOR FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: US 60/312,726  
PRIOR FILING DATE: 2001-08-16  
NUMBER OF SEQ ID NOS: 23  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 56  
TYPE: PRT  
ORGANISM: Homo sapiens  
-10-219-329-4

Query Match 100.0%; Score 19; DB 14; Length 56;  
Best Local Similarity 100.0%; Pred. No. 7.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||||  
31 VAEF 34

SULT 31  
-10-153-185-4  
Sequence 4, Application US/10153185  
Publication No. US20030148959A1  
GENERAL INFORMATION:  
APPLICANT: Quirk, Stephen  
APPLICANT: Malik, Sohail  
TITLE OF INVENTION: Anti-Aging and Wound Healing Compounds  
FILE REFERENCE: 1443.034U01  
CURRENT APPLICATION NUMBER: US/10/153,185  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US 10/032,376  
PRIOR FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: US 60/312,726  
PRIOR FILING DATE: 2001-08-16  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 56  
TYPE: PRT  
ORGANISM: Homo sapiens  
-10-153-185-4

Query Match 100.0%; Score 19; DB 14; Length 56;  
Best Local Similarity 100.0%; Pred. No. 7.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||||  
31 VAEF 34

SULT 32  
-10-219-561-4

Sequence 4, Application US/10219561  
Publication No. US20030166567A1  
GENERAL INFORMATION:  
APPLICANT: Quirk, Stephen  
APPLICANT: Malik, Sohail  
APPLICANT: Villanueva, Julie M.  
TITLE OF INVENTION: Anti-Aging and Wound Healing Compounds  
FILE REFERENCE: 1443.008U02  
CURRENT APPLICATION NUMBER: US/10/219,561  
CURRENT FILING DATE: 2002-08-15  
PRIOR APPLICATION NUMBER: US 10/032,376  
PRIOR FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: US 10/153,185  
PRIOR FILING DATE: 2002-05-21  
PRIOR APPLICATION NUMBER: US 60/312,726  
PRIOR FILING DATE: 2001-08-16  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 56  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-219-561-4

Query Match 100.0%; Score 19; DB 14; Length 56;  
Best Local Similarity 100.0%; Pred. No. 7.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||||  
31 VAEF 34

RESULT 33  
US-10-424-599-215734  
Sequence 215734, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 215734  
LENGTH: 57  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_36836C.1.psp  
US-10-424-599-215734

Query Match 100.0%; Score 19; DB 12; Length 57;  
Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||||  
48 VAEF 51

RESULT 34  
US-10-424-599-260762  
Sequence 260762, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua

APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 260762  
LENGTH: 57  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_77490C.1.pcp  
S-10-424-599-260762

Query Match 100.0%; Score 19; DB 12; Length 57;  
Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 VAEF 4  
b 4 VAEF 7

RESULT 35  
S-10-372-003A-75  
Sequence 75, Application US/10372003A  
Publication No. US20030215846A1  
GENERAL INFORMATION:  
APPLICANT: Thomas, Wayne  
APPLICANT: Watt, Paul  
APPLICANT: Hopkins, Richard  
TITLE OF INVENTION: Methods of constructing and screening  
FILE OF INVENTION: diverse expression libraries  
FILE REFERENCE: PERIC40.001CPI  
CURRENT APPLICATION NUMBER: US/10/372,003A  
CURRENT FILING DATE: 2003-02-21  
PRIOR APPLICATION NUMBER: US 09/568,229  
PRIOR FILING DATE: 2000-05-05  
PRIOR APPLICATION NUMBER: US 60/132,711  
PRIOR FILING DATE: 1999-05-05  
NUMBER OF SEQ ID NOS: 81  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 75  
LENGTH: 57  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: synthetic clone BGF24 with first open reading  
S-10-372-003A-75

Query Match 100.0%; Score 19; DB 15; Length 57;  
Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 VAEF 4  
b 28 VAEF 31

RESULT 36  
S-10-424-599-214930  
Sequence 214930, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B

CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 214930  
LENGTH: 58  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_36107C.1.pcp  
US-10-424-599-214930

Query Match 100.0%; Score 19; DB 12; Length 58;  
Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 13 VAEF 16

RESULT 37  
US-09-864-761-45291  
Sequence 45291, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wenheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/006666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 45291

```
LENGTH: 60
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC009079.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.54
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.77
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.65
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.74
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.48
OTHER INFORMATION: EST HUMAN HIT: AV756022.1, EVALUATE 6.00e-28
OTHER INFORMATION: SWISSPROT HIT: P56315, EVALUATE 2.00e-00
-09-864-761-45291

Query Match      100.0%; Score 19; DB 9; Length 60;
Best Local Similarity 100.0%; Pred. No. 9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4
|||||
8 VAEF 11

SULT 38
-09-393-634-80
Sequence 80, Application US/09393634
Patent No. US20020051997A1
GENERAL INFORMATION:
APPLICANT: Zuker, Charles S.
APPLICANT: Adler, Jon Elliot
APPLICANT: Ryba, Nick
APPLICANT: Mueller, Ken
APPLICANT: Hoon, Mark
APPLICANT: The Regents of the University of California
APPLICANT: The Government of the United States of America
APPLICANT: as represented by the Secretary of the
APPLICANT: Department of Health and Human Services
TITLE OF INVENTION: SF, a No. US20020051997A1el Family of Taste Receptors
FILE REFERENCE: 02307E-098000US
CURRENT APPLICATION NUMBER: US/09/393,634
CURRENT FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 80
LENGTH: 68
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human GR24
-09-393-634-80

Query Match      100.0%; Score 19; DB 9; Length 68;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4
|||||
14 VAEF 17

SULT 39
-09-864-408A-40
Sequence 40, Application US/09864408A
Publication No. US20040009474A1
GENERAL INFORMATION:
APPLICANT: Leach, Martin D.
APPLICANT: Shinkets, Richard A.
TITLE OF INVENTION: No. US20040009474A1el Human Polynucleotides and Polypeptides Enco
FILE REFERENCE: 21402-012
CURRENT APPLICATION NUMBER: US/09/864,408A
CURRENT FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: 60/206,690
PRIOR FILING DATE: 2000-05-24

; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-40

Query Match      100.0%; Score 19; DB 11; Length 68;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4
|||||
3 VAEF 6

RESULT 40
US-10-364-861-80
Sequence 80, Application US/10364861
Publication No. US20040038312A1
GENERAL INFORMATION:
APPLICANT: Zuker, Charles S.
APPLICANT: Adler, Jon Elliot
APPLICANT: Hoon, Mark
APPLICANT: Ryba, Nick
APPLICANT: Mueller, Ken
APPLICANT: The Regents of the University of California
APPLICANT: The Government of the United States of America
APPLICANT: as represented by the Secretary of the
APPLICANT: Department of Health and Human Services
TITLE OF INVENTION: T2R, a No. US20040038312A1el Family of Taste Receptors
FILE REFERENCE: 02307E-098020US
CURRENT APPLICATION NUMBER: US/10/364,861
CURRENT FILING DATE: 2003-06-30
PRIOR APPLICATION NUMBER: US 09/393,634
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 80
LENGTH: 68
TYPE: PRT
FEATURE:
OTHER INFORMATION: human T2R24, GR24 or SF24
US-10-364-861-80

Query Match      100.0%; Score 19; DB 12; Length 68;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4
|||||
14 VAEF 17

RESULT 41
US-10-383-982-80
Sequence 80, Application US/10383982
Publication No. US20030157568A1
GENERAL INFORMATION:
APPLICANT: Zuker, Charles S.
APPLICANT: Adler, Jon Elliot
APPLICANT: Ryba, Nick
APPLICANT: Mueller, Ken
APPLICANT: Hoon, Mark
APPLICANT: The Regents of the University of California
APPLICANT: The Government of the United States of America
APPLICANT: as represented by the Secretary of the
APPLICANT: Department of Health and Human Services
TITLE OF INVENTION: SF, a No. US20030157568A1el Family of Taste Receptors
FILE REFERENCE: 02307E-098000US
CURRENT APPLICATION NUMBER: US/10/383,982
```

CURRENT FILING DATE: 2003-03-07  
PRIOR APPLICATION NUMBER: US/09/393,634  
PRIOR FILING DATE: 1999-09-10  
NUMBER OF SEQ ID NOS: 92  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 80  
LENGTH: 68  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: human GR24  
;-10-383-982-80

Query Match 100.0%; Score 19; DB 14; Length 68;  
Best Local Similarity 100.0%; Pred. No. 9.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||  
14 VAEF 17

RESULT 42  
;-09-864-761-45716  
Sequence 45716, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn. Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aeonica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24363.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 45716  
LENGTH: 69  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL162171.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.81  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.91  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.85  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.72  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.71  
OTHER INFORMATION: EST HUMAN HIT: BE55987.1, EVALUE 2.00e-03  
OTHER INFORMATION: SWISSPROT HIT: P56093, EVALUE 2.00e+00  
US-09-864-761-45716

Query Match 100.0%; Score 19; DB 9; Length 69;  
Best Local Similarity 100.0%; Pred. No. 9.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||  
15 VAEF 18

RESULT 43  
US-09-738-626-5596  
Sequence 5596, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NACKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OKAZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280989  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 5596  
LENGTH: 69  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5596

Query Match 100.0%; Score 19; DB 9; Length 69;  
Best Local Similarity 100.0%; Pred. No. 9.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||  
23 VAEF 26

RESULT 44  
US-10-424-599-185385  
Sequence 185385, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J

```

APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 185385
LENGTH: 69
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_138417C.1.pep
-10-424-599-185385

Query Match 100.0%; Score 19; DB 12; Length 69;
Best Local Similarity 100.0%; Pred. No. 9.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4
|||||
24 VAEF 27

SULT 45
-10-029-386-28645
Sequence 28645, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
FILE REFERENCE: AEOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2003-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 28645
LENGTH: 70
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO CHR17.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.8
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 8.7
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7
OTHER INFORMATION: SWISSPROT HIT: P78716, EVALUATION 3.00e-04
-10-029-386-28645

Query Match 100.0%; Score 19; DB 14; Length 70;
Best Local Similarity 100.0%; Pred. No. 9.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4
|||||
43 VAEF 46

SULT 46
-10-424-599-202605
Sequence 202605, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J

APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 202605
LENGTH: 71
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_24978C.1.pep
US-10-424-599-202605

Query Match 100.0%; Score 19; DB 12; Length 71;
Best Local Similarity 100.0%; Pred. No. 9.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4
|||||
17 VAEF 20

RESULT 47
US-10-424-599-274981
Sequence 274981, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 274981
LENGTH: 71
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(71)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_90329C.1.pep
US-10-424-599-274981

Query Match 100.0%; Score 19; DB 12; Length 71;
Best Local Similarity 100.0%; Pred. No. 9.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4
|||||
39 VAEF 42

RESULT 48
US-10-029-386-30348
Sequence 30348, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
FILE REFERENCE: AEOMICA-X-2

```

CURRENT APPLICATION NUMBER: US/10/029,386

CURRENT FILING DATE: 2001-12-20

NUMBER OF SEQ ID NOS: 34288

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 30348

LENGTH: 71

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO CHR7.1

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.82

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.73

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 6.2

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1

OTHER INFORMATION: SWISSPROT HIT: P21333, EVALUATION 5.00e-22

S-10-029-386-30348

Query Match 100.0%; Score 19; DB 14; Length 71;

Best Local Similarity 100.0%; Pred. No. 9.5e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 VAEF 4

b 60 VAEF 63

RESULT 49

S-10-029-386-29320

Sequence 29320, Application US/10029386

Publication No. US20030194704A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: AEMICA-X-2

CURRENT FILING DATE: 2001-12-20

CURRENT APPLICATION NUMBER: US/10/029,386

NUMBER OF SEQ ID NOS: 34288

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 29320

LENGTH: 72

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO CHR2.1

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.99

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.6

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.4

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.56

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.81

OTHER INFORMATION: SWISSPROT HIT: P14650, EVALUATION 6.00e-14

S-10-029-386-29320

Query Match 100.0%; Score 19; DB 14; Length 72;

Best Local Similarity 100.0%; Pred. No. 9.6e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 VAEF 4

b 45 VAEF 48

RESULT 50

S-10-214-188-10

Sequence 10, Application US/10214188

Publication No. US2003002260A1

GENERAL INFORMATION:

APPLICANT: LA THANGUE, NICHOLAS B.

BERNARDS, RENE  
HILJANS, ELEANORE M.  
TITLE OF INVENTION: TRANSCRIPTION FACTOR E2F-5  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHUYE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/214,188  
FILING DATE: 08-Aug-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,139  
FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: WILSON, MARY J.  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 620-22  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 74 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-10-214-188-10

Query Match 100.0%; Score 19; DB 14; Length 74;

Best Local Similarity 100.0%; Pred. No. 9.9e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4

Db 23 VAEF 26

RESULT 51

US-10-156-761-9071

Sequence 9071, Application US/10156761

Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 9071

LENGTH: 74

TYPE: PRT

ORGANISM: Streptomyces avermitilis  
-10-156-761-9071

Query Match 100.0%; Score 19; DB 14; Length 74;  
Best Local Similarity 100.0%; Pred. No. 9.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
|||  
36 VAEF 39

SULT 52  
-10-424-599-156128  
Sequence 156128, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 156128  
LENGTH: 75  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(75)  
OTHER INFORMATION: unsure at all Xaa locations  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_112004C.1.pap  
-10-424-599-156128

Query Match 100.0%; Score 19; DB 12; Length 75;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4  
|||  
2 VAEF 5

SULT 53  
-10-424-599-261424  
Sequence 261424, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 261424  
LENGTH: 75  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_78089C.1.pap  
-10-424-599-261424

Query Match 100.0%; Score 19; DB 12; Length 75;  
Best Local Similarity 100.0%; Pred. No. 1e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
|||  
DB 12 VAEF 15

RESULT 54  
US-10-424-599-253204  
Sequence 253204, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 253204  
LENGTH: 76  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(76)  
OTHER INFORMATION: unsure at all Xaa locations  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_70669C.1.pap  
US-10-424-599-253204

Query Match 100.0%; Score 19; DB 12; Length 76;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||  
DB 64 VAEF 67

RESULT 55  
US-10-424-599-211908  
Sequence 211908, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 211908  
LENGTH: 77  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_3337C.1.pap  
US-10-424-599-211908

Query Match 100.0%; Score 19; DB 12; Length 77;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
|||  
DB 16 VAEF 19

RESULT 56  
S-10-335-977-7493  
Sequence 7493, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 7493:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 77 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...77  
SEQUENCE DESCRIPTION: SEQ ID NO: 7493:  
S-10-335-977-7493  
Query Match 100.0%; Score 19; DB 12; Length 77;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 VAEF 4  
|||  
b 40 VAEF 43  
RESULT 57  
US-10-029-386-32709  
Sequence 32709, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
FILE REFERENCE: AEOMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386

CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 32709  
LENGTH: 77  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC004186.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.76  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5  
OTHER INFORMATION: SWISSPROT HIT: P14373, EVALUATE 7.50e+00  
US-10-029-386-32709  
Query Match 100.0%; Score 19; DB 14; Length 77;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
|||  
Db 11 VAEF 14  
RESULT 58  
US-10-424-599-162039  
Sequence 162039, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 162039  
LENGTH: 78  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_117339C.1.pep  
US-10-424-599-162039  
Query Match 100.0%; Score 19; DB 12; Length 78;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
|||  
Db 28 VAEF 31  
RESULT 59  
US-10-424-599-262701  
Sequence 262701, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 262701  
LENGTH: 78

```
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_79240C.1.pap
-10-424-599-262701

Query Match      100.0%; Score 19; DB 12; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

      1 VAEF 4
      |||||
      5 VAEF 8

SULT 60
-10-363-616-368
Sequence 368, Application US/10363616
Publication No. US20040044181A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 21272-113 (793)
CURRENT APPLICATION NUMBER: US/10/363,616
CURRENT FILING DATE: 2003-03-03
PRIOR APPLICATION NUMBER: 09/654,935
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 490
SEQ ID NO 368
LENGTH: 79
TYPE: PRT
ORGANISM: Homo sapiens
-10-363-616-368

Query Match      100.0%; Score 19; DB 12; Length 79;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

      1 VAEF 4
      |||||
      24 VAEF 27

SULT 61
-10-424-599-216097
Sequence 216097, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 216097
LENGTH: 80
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_37165C.1.pap
-10-424-599-216097

Query Match      100.0%; Score 19; DB 12; Length 80;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

      1 VAEF 4
      |||||
      35 VAEF 38

RESULT 62
US-10-367-980A-14
; Sequence 14, Application US/10367980A
; Publication No. US20030228592A1
; GENERAL INFORMATION:
; APPLICANT: St Vincent's Institute of Medical Research
; APPLICANT: Rogers, Suzanne D
; APPLICANT: Best, James D
; TITLE OF INVENTION: Human Facilitative Glucose Transport Protein GLUT2
; FILE REFERENCE: VS-AJH:FP17928
; CURRENT APPLICATION NUMBER: US/10/367,980A
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: 09/509,731
; PRIOR FILING DATE: 2003-02-19
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 14
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GLUT2
US-10-367-980A-14

Query Match      100.0%; Score 19; DB 15; Length 82;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

      1 VAEF 4
      |||||
      13 VAEF 16

RESULT 63
US-09-864-408A-5282
; Sequence 5282, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5282
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-5282

Query Match      100.0%; Score 19; DB 11; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

      1 VAEF 4
      |||||
      52 VAEF 55

RESULT 64
US-10-078-090-125
; Sequence 125, Application US/10078090
; Publication No. US20030044815A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
```

APPLICANT: Hu, Ping  
APPLICANT: Recipon, Herve  
APPLICANT: Karra, Kalpana  
APPLICANT: Caferkey, Robert  
APPLICANT: Sun, Yongming  
APPLICANT: Liu, Chenghua  
TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr  
FILE REFERENCE: DEX-0312  
CURRENT APPLICATION NUMBER: US/10/078,090  
CURRENT FILING DATE: 2002-02-14  
PRIOR APPLICATION NUMBER: 60/268,999  
PRIOR FILING DATE: 2001-02-15  
NUMBER OF SEQ ID NOS: 210  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 125  
LENGTH: 84  
TYPE: PRT  
ORGANISM: Homo sapien  
S-10-078-090-125  
Query Match 100.0%; Score 19; DB 14; Length 84;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 VAEF 4  
b 38 VAEF 41  
ESULT 65  
S-10-424-599-155636  
Sequence 155636, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Zhou Yihua  
APPLICANT: Kovalic David K  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 155636  
LENGTH: 85  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_11155C.1.pep  
S-10-424-599-155636  
Query Match 100.0%; Score 19; DB 12; Length 85;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 VAEF 4  
b 46 VAEF 49  
ESULT 66  
S-10-424-599-201853  
Sequence 201853, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Zhou Yihua  
APPLICANT: Kovalic David K  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B

APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11224  
LENGTH: 89  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
S-09-815-242-11224

Query Match 100.0%; Score 19; DB 12; Length 85;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 58 VAEF 61

## RESULT 67

US-09-815-242-11224  
Sequence 11224, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11224  
LENGTH: 89  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
S-09-815-242-11224

Query Match 100.0%; Score 19; DB 9; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4  
Db 12 VAEF 15

SUIT 68  
 -09-815-242-11245  
 Sequence 11245, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kari L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: Identification of Essential Genes in  
 TITLE OF INVENTION: Prokaryotes  
 FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 CURRENT FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/267,636  
 PRIOR FILING DATE: 2001-02-09  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 Remaining Prior Application data removed - See File Wrapper or PALM.  
 NUMBER OF SEQ ID NOS: 78614  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 58489  
 LENGTH: 89  
 TYPE: PRT  
 ORGANISM: Haemophilus influenzae  
 US-10-282-122A-58489  
 Query Match 100.0%; Score 19; DB 12; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 VAEF 4  
 Db 12 VAEF 15  
 RESULT 70  
 US-10-282-122A-58530  
 Sequence 58530, Application US/10282122A  
 Publication No. US20040029129A1  
 GENERAL INFORMATION:  
 APPLICANT: Wang, Liangsu  
 APPLICANT: Zamudio, Carlos  
 APPLICANT: Malone, Cheryl  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kari  
 APPLICANT: Zyskind, Judith  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John  
 APPLICANT: Carr, Grant  
 APPLICANT: Yamamoto, Robert  
 APPLICANT: Forsyth, R.  
 APPLICANT: Xu, H.  
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 FILE REFERENCE: ELITRA.034A  
 CURRENT APPLICATION NUMBER: US/10/282,122A  
 CURRENT FILING DATE: 2003-02-20  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/230,335  
 PRIOR FILING DATE: 2000-09-06  
 PRIOR APPLICATION NUMBER: 60/230,347  
 PRIOR FILING DATE: 2000-09-09  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/267,636

SUIT 68  
 -09-815-242-11245  
 Sequence 11245, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kari L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: Identification of Essential Genes in  
 TITLE OF INVENTION: Prokaryotes  
 FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 CURRENT FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/267,636  
 PRIOR FILING DATE: 2001-02-09  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 Remaining Prior Application data removed - See File Wrapper or PALM.  
 NUMBER OF SEQ ID NOS: 78614  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 58489  
 LENGTH: 89  
 TYPE: PRT  
 ORGANISM: Haemophilus influenzae  
 US-10-282-122A-58489  
 Query Match 100.0%; Score 19; DB 12; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 VAEF 4  
 Db 12 VAEF 15  
 RESULT 70  
 US-10-282-122A-58530  
 Sequence 58530, Application US/10282122A  
 Publication No. US20040029129A1  
 GENERAL INFORMATION:  
 APPLICANT: Wang, Liangsu  
 APPLICANT: Zamudio, Carlos  
 APPLICANT: Malone, Cheryl  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kari  
 APPLICANT: Zyskind, Judith  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John  
 APPLICANT: Carr, Grant  
 APPLICANT: Yamamoto, Robert  
 APPLICANT: Forsyth, R.  
 APPLICANT: Xu, H.  
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 FILE REFERENCE: ELITRA.034A  
 CURRENT APPLICATION NUMBER: US/10/282,122A  
 CURRENT FILING DATE: 2003-02-20  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/230,335  
 PRIOR FILING DATE: 2000-09-06  
 PRIOR APPLICATION NUMBER: 60/230,347  
 PRIOR FILING DATE: 2000-09-09  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 58530  
LENGTH: 89  
TYPE: PRT  
ORGANISM: Haemophilus influenzae

S-10-282-122A-58530

Query Match 100.0%; Score 19; DB 12; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4  
b 12 VAEF 15

RESULT 71

S-10-282-122A-66960  
Sequence 66960, Application US/10282122A  
Publication No. US20040029129A1

GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 66960

LENGTH: 89

TYPE: PRT

ORGANISM: Pasteurella multocida

S-10-282-122A-66960

Query Match 100.0%; Score 19; DB 12; Length 89;

Best Local Similarity 100.0%; Pred. No. 1.2e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VAEF 4

b 12 VAEF 15

RESULT 73

S-10-282-122A-599-180307

Sequence 180307, Application US/10424599

Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J

APPLICANT: Kovalic David K

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VAEF 4  
Db 12 VAEF 15

RESULT 72

US-10-282-122A-68774  
Sequence 68774, Application US/10282122A  
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 68774

LENGTH: 89

TYPE: PRT

ORGANISM: Proteus mirabilis

US-10-282-122A-68774

Query Match 100.0%; Score 19; DB 12; Length 89;

Best Local Similarity 100.0%; Pred. No. 1.2e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4

Db 12 VAEF 15

RESULT 73

US-10-424-599-180307

Sequence 180307, Application US/10424599

Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J

APPLICANT: Kovalic David K

APPLICANT: Zhou Yihua  
 APPLICANT: Cao Yongwei  
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 Title of Invention: Plants and Uses Thereof for Plant Improvement  
 FILE REFERENCE: 38-21(53223)B  
 CURRENT APPLICATION NUMBER: US/10/424,599  
 CURRENT FILING DATE: 2003-04-28  
 NUMBER OF SEQ ID NOS: 285684

SEQ ID NO 180307

LENGTH: 89

TYPE: PRT

ORGANISM: Glycine max

FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MRT3847\_133832C.1.pgp

-10-424-599-180307

Query Match 100.0%; Score 19; DB 12; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4

24 VAEF 27

SULT 74

-10-097-111-360

Sequence 360, Application US/10097111

Publication No. US2003013877A1

GENERAL INFORMATION:

APPLICANT: PELLETIER, JERRY

APPLICANT: GROS, PHILLIPPE

APPLICANT: DUBOW, MICHAEL

TITLE OF INVENTION: DNA SEQUENCES FROM S. PNEUMONIAE BACTERIOPHAGE EPI THAT

FILE REFERENCE: ENCODE ANTI-MICROBIAL POLYPEPTIDES

FILE REFERENCE: 073406-0503

CURRENT APPLICATION NUMBER: US/10/097,111

CURRENT FILING DATE: 2002-07-24

PRIOR APPLICATION NUMBER: 09/676,412

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/157,218

PRIOR FILING DATE: 1999-09-30

NUMBER OF SEQ ID NOS: 552

SOFTWARE: Patent in Ver. 2.1

SEQ ID NO 360

LENGTH: 89

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

-10-097-111-360

Query Match 100.0%; Score 19; DB 14; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VAEF 4

11 VAEF 14

SULT 75

-10-424-599-245830

Sequence 245830, Application US/10424599

Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J

APPLICANT: Kowalic David K

APPLICANT: Zhou Yihua

APPLICANT: Cao Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

Title of Invention: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53223)B

CURRENT APPLICATION NUMBER: US/10/424,599

CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 245830  
 ; LENGTH: 90  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_64015C.1.pgp  
 US-10-424-599-245830

Query Match 100.0%; Score 19; DB 12; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAEF 4

Db 60 VAEF 63

Search completed: May 24, 2004, 17:42:33  
 Job time : 32.7143 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on on: May 24, 2004, 17:31:37 ; Search time 51.7857 Seconds

(without alignments)

27.280 Million cell updates/sec

File: US-09-594-978a-3

irect score: 20

quence: 1 XVAEP 5

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

ital number of hits satisfying chosen parameters: 1586107

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : A\_Geneseq\_29Jan04.\*

1: Geneseq1980s.\*

2: Geneseq1990s.\*

3: Geneseq2000s.\*

4: Geneseq2001s.\*

5: Geneseq2002s.\*

6: Geneseq2003as.\*

7: Geneseq2003bs.\*

8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	95.0	6	5	AAM47151 S chryson
2	19	95.0	7	3	AB07871 A beta-se
3	19	95.0	8	3	AB07872 A beta-se
4	19	95.0	8	4	AB056241 Vascular
5	19	95.0	8	4	AAU28720 DFI trypt
6	19	95.0	8	4	AAU25114 Schizophr
7	19	95.0	8	4	AAU26368 Depressio
8	19	95.0	8	4	AAU15458 Schizophr
9	19	95.0	8	2	AAU78909 MAGE 3 10
10	19	95.0	9	5	AB07871 A beta-am
11	19	95.0	10	5	AB06593 Beta-secr
12	19	95.0	12	5	AB06592 Beta-secr
13	19	95.0	12	5	ARG78404 Memapsin
14	19	95.0	12	7	ADD35467 Escherich
15	19	95.0	13	3	AB07889 A beta-se
16	19	95.0	13	4	AAU92776 Vaccine r
17	19	95.0	14	3	AB07888 A peptide
18	19	95.0	14	4	AAU97460 Human pep
19	19	95.0	15	6	AAE32223 Arabidops
20	19	95.0	20	2	AAU42944 Immunogen
21	19	95.0	20	2	AAU42943 Immunogen
22	19	95.0	21	4	AB069462 Synthetic
23	19	95.0	21	4	AB069463 Synthetic
24	19	95.0	25	2	AAU42946 Immunogen
25	19	95.0	25	4	AAU69465 Synthetic

26	19	95.0	28	2	AAW17832	Human mev
27	19	95.0	28	4	AAW33910	Peptide #
28	19	95.0	28	4	ABG55464	Human liv
29	19	95.0	28	5	ABG33601	Human pep
30	19	95.0	30	2	AAW62760	Streptococ
31	19	95.0	32	4	AAW82654	Human imm
32	19	95.0	33	4	AAO09024	Human pol
33	19	95.0	33	2	AAW77953	Antigenic
34	19	95.0	35	2	AAW46146	Predicted
35	19	95.0	35	2	AAW51751	H. influe
36	19	95.0	35	2	AAW55788	Human aqu
37	19	95.0	35	2	AAW54105	Tdp2 anti
38	19	95.0	35	2	AAW94320	Human aqu
39	19	95.0	35	3	AAW80448	H. influe
40	19	95.0	36	2	AAW27782	UDP-N-ace
41	19	95.0	37	4	AAW99833	Human exc
42	19	95.0	37	4	AAW42648	Human kid
43	19	95.0	37	4	AAW17544	Peptide #
44	19	95.0	43	4	ABB31362	Peptide #
45	19	95.0	43	4	ABB21904	Protein #
46	19	95.0	43	4	ABG51414	Human liv
47	19	95.0	43	4	AAW05214	Peptide #
48	19	95.0	47	3	AAW27178	RSV parti
49	19	95.0	48	4	AAW18541	Peptide #
50	19	95.0	48	4	ABB37578	Peptide #
51	19	95.0	48	4	AAW30996	Peptide #
52	19	95.0	48	4	ABB22868	Protein #
53	19	95.0	48	4	AAW70687	Human bon
54	19	95.0	48	4	ABG52389	Human liv
55	19	95.0	48	4	AAW06107	Peptide #
56	19	95.0	53	4	AAO10565	Human pol
57	19	95.0	53	4	AAO05247	Human pol
58	19	95.0	56	6	AAW87390	Human imm
59	19	95.0	56	6	ABP97126	Human mat
60	19	95.0	56	6	ABG76312	Human mat
61	19	95.0	58	2	AAW55888	Rat PC12
62	19	95.0	58	5	ABP04854	Human ORF
63	19	95.0	59	3	AAW60085	Arabidops
64	19	95.0	59	3	AAW60683	Arabidops
65	19	95.0	60	4	ABB42044	Peptide #
66	19	95.0	60	4	AAW35846	Peptide #
67	19	95.0	60	4	AAW75737	Human bon
68	19	95.0	60	4	AAW62925	Human bra
69	19	95.0	60	4	ABG57475	Human liv
70	19	95.0	62	2	AAW17262	HLH Domai
71	19	95.0	62	4	AAW51247	Human imm
72	19	95.0	62	6	AAW65667	Protonib
73	19	95.0	64	4	AAO09517	Human pol
74	19	95.0	67	2	AAW48352	Human pro
75	19	95.0	68	5	ABP31047	Human ORF
76	19	95.0	69	3	AAW32809	Zea ways
77	19	95.0	69	3	AAW21161	Zea ways
78	19	95.0	69	4	ABB42991	Peptide #
79	19	95.0	69	4	AAW36821	Peptide #
80	19	95.0	69	4	AAW76713	Human bon
81	19	95.0	69	4	AAW63900	Human bra
82	19	95.0	69	4	ABG58405	Human liv
83	19	95.0	69	4	AAW31842	C glutami
84	19	95.0	71	5	ABP08064	Hepatitis
85	19	95.0	71	5	AAW15877	Hepatitis
86	19	95.0	71	6	AAW35824	Acinetoba
87	19	95.0	73	4	AAW33554	Protonib
88	19	95.0	73	4	AAW66465	Protonib
89	19	95.0	73	5	ABP06455	Human ORF
90	19	95.0	73	6	ABM50073	Protonib
91	19	95.0	73	6	ABM62984	Protonib
92	19	95.0	74	5	AAW15876	Hepatitis
93	19	95.0	75	4	ABG00769	Novel hum
94	19	95.0	76	4	AAW45469	Protonib
95	19	95.0	76	5	ABM41988	Protonib
96	19	95.0	79	5	ABP62931	Human pol
97	19	95.0	83	6	ABP79418	N. gonorr
98	19	95.0	84	5	ABP66623	Human bre

99	19	95.0	84	5	ABP33668	Human ORF	172	19	95.0	133	4	ABB27760	Human pep
100	19	95.0	86	4	AAU50485	Propionib	173	19	95.0	133	4	ABB18410	Propionib
101	19	95.0	86	6	ABM47004	Propionib	174	19	95.0	133	4	AAAM66115	Human bon
102	19	95.0	87	4	AAOC5209	Human pol	175	19	95.0	133	4	AAAM53732	Human bra
103	19	95.0	87	4	AAU43264	Propionib	176	19	95.0	133	4	ABG47783	Human liv
104	19	95.0	87	4	AAU66125	Propionib	177	19	95.0	133	4	AAAM01727	Peptide #
105	19	95.0	87	6	ABM62644	Propionib	178	19	95.0	133	5	ABP10781	Human ORF
106	19	95.0	87	6	ABM39783	Propionib	179	19	95.0	133	5	ABG35765	Human pep
107	19	95.0	88	4	AAU56968	Propionib	180	19	95.0	134	4	AAU34920	Enterococ
108	19	95.0	88	6	ABM53487	Propionib	181	19	95.0	134	4	ABG02560	Novel hum
109	19	95.0	88	7	ABM53487	Propionib	182	19	95.0	135	5	ABP07711	Human ORF
110	19	95.0	89	3	ABM16761	Bacteriop	183	19	95.0	135	6	ABM69996	Phototrab
111	19	95.0	89	4	AAU35652	Haemophil	184	19	95.0	136	2	AAW55889	Rat plas
112	19	95.0	89	4	AAU35631	Haemophil	185	19	95.0	139	3	AAAB10817	Mechanoba
113	19	95.0	89	5	ABM54895	Lactococc	186	19	95.0	139	3	AAAB1153	Zea may
114	19	95.0	89	6	ABU40850	Protein e	187	19	95.0	141	3	AAAG02426	Human aci
115	19	95.0	89	6	ABU30565	Protein e	188	19	95.0	141	3	AAAG02426	Human sec
116	19	95.0	89	6	ABU30606	Protein e	189	19	95.0	141	4	AAU50269	Propionib
117	19	95.0	89	6	ABU30606	Protein e	190	19	95.0	141	4	ABG30243	Novel hum
118	19	95.0	89	6	ABU30936	Protein e	191	19	95.0	141	4	ABG07245	Novel hum
119	19	95.0	93	3	ACG48060	Arabidops	192	19	95.0	141	6	ABM46788	Propionib
120	19	95.0	96	3	ACG34249	Arabidops	193	19	95.0	141	6	ABM46788	Propionib
121	19	95.0	96	3	ABG07485	Novel hum	194	19	95.0	142	6	ABP73097	Canola co
122	19	95.0	97	5	ABM48133	Listeria	195	19	95.0	142	6	ABP73097	N. gonorr
123	19	95.0	97	5	ABP75586	Human sec	196	19	95.0	143	3	AAAB10816	Archaeogl
124	19	95.0	99	5	ABP41643	Human ova	197	19	95.0	143	5	AAO14828	Human ste
125	19	95.0	100	3	ACG02653	Propionib	198	19	95.0	144	3	AAAG05007	Arabidops
126	19	95.0	100	6	ABM58784	Propionib	199	19	95.0	145	2	AAAY34749	Chlamydia
127	19	95.0	105	3	ACG32808	Zea may	200	19	95.0	145	2	AAU27872	Human con
128	19	95.0	106	3	ACG32572	Arabidops	201	19	95.0	145	5	ABG98409	Haicarcu
129	19	95.0	106	5	ABM82190	P. multoc	202	19	95.0	145	5	ABG98409	Haicarcu
130	19	95.0	107	3	ACG61315	Arabidops	203	19	95.0	145	5	ABJ11020	Yeast sel
131	19	95.0	107	3	ACG10318	Arabidops	204	19	95.0	145	5	ABP43835	Yeast sel
132	19	95.0	107	3	ACG0766	C glutami	205	19	95.0	146	6	ABP57363	RIKEN 170
133	19	95.0	108	3	ACG00023	Human sec	206	19	95.0	146	6	ABU42336	Protein e
134	19	95.0	109	4	ABG09614	Novel hum	207	19	95.0	146	7	ADP26750	Human adi
135	19	95.0	109	5	ABM89162	Human pol	208	19	95.0	147	6	ABU07050	Maize SSI
136	19	95.0	109	5	ABP62930	Human pol	209	19	95.0	148	3	ACG41152	Zea may
137	19	95.0	109	5	ABP62930	Human pol	210	19	95.0	148	6	ABP78900	N. gonorr
138	19	95.0	111	2	AAV37827	Amino aci	211	19	95.0	148	7	ABM74182	DNA clone
139	19	95.0	111	2	AAV37827	Amino aci	212	19	95.0	148	7	ABM74182	DNA clone
140	19	95.0	112	1	AAV37827	Amino aci	213	19	95.0	149	5	ABG05510	Novel hum
141	19	95.0	112	1	AAV37827	Amino aci	214	19	95.0	149	5	ABG05510	Novel hum
142	19	95.0	113	1	AAV37827	Amino aci	215	19	95.0	149	5	ABG05510	Novel hum
143	19	95.0	113	3	ACG34248	Arabidops	216	19	95.0	149	5	ABG05510	Novel hum
144	19	95.0	113	3	ACG34248	Arabidops	217	19	95.0	149	5	ABG05510	Novel hum
145	19	95.0	114	1	AAV37827	Amino aci	218	19	95.0	149	5	ABG05510	Novel hum
146	19	95.0	114	6	ABP97830	Human RS3	219	19	95.0	150	3	ABG07294	Novel hum
147	19	95.0	115	5	ABM62629	Propionib	220	19	95.0	150	3	ABG07294	Novel hum
148	19	95.0	116	6	ABM64780	Propionib	221	19	95.0	150	3	ABG07294	Novel hum
149	19	95.0	117	5	ABP05042	Human ORF	222	19	95.0	150	3	ABG07294	Novel hum
150	19	95.0	119	3	ABM33108	Pinus rad	223	19	95.0	150	3	ABG07294	Novel hum
151	19	95.0	119	4	ABM69462	Drosophil	224	19	95.0	150	3	ABG07294	Novel hum
152	19	95.0	119	6	ABP70942	Drosophil	225	19	95.0	150	3	ABG07294	Novel hum
153	19	95.0	122	4	ABM16105	Human rer	226	19	95.0	150	3	ABG07294	Novel hum
154	19	95.0	122	5	ABM16105	Human rer	227	19	95.0	150	3	ABG07294	Novel hum
155	19	95.0	123	4	AAV31290	Human ORF	228	19	95.0	150	3	ABG07294	Novel hum
156	19	95.0	123	4	AAV31290	Human ORF	229	19	95.0	150	3	ABG07294	Novel hum
157	19	95.0	123	4	AAU67165	Propionib	230	19	95.0	150	3	ABG07294	Novel hum
158	19	95.0	123	6	ABM63684	Propionib	231	19	95.0	150	3	ABG07294	Novel hum
159	19	95.0	123	6	ABM63684	Propionib	232	19	95.0	150	3	ABG07294	Novel hum
160	19	95.0	123	6	ABM63684	Propionib	233	19	95.0	150	3	ABG07294	Novel hum
161	19	95.0	123	6	ABM63684	Propionib	234	19	95.0	150	3	ABG07294	Novel hum
162	19	95.0	123	6	ABM63684	Propionib	235	19	95.0	150	3	ABG07294	Novel hum
163	19	95.0	123	6	ABM63684	Propionib	236	19	95.0	150	3	ABG07294	Novel hum
164	19	95.0	123	6	ABM63684	Propionib	237	19	95.0	150	3	ABG07294	Novel hum
165	19	95.0	123	6	ABM63684	Propionib	238	19	95.0	150	3	ABG07294	Novel hum
166	19	95.0	123	6	ABM63684	Propionib	239	19	95.0	150	3	ABG07294	Novel hum
167	19	95.0	123	6	ABM63684	Propionib	240	19	95.0	150	3	ABG07294	Novel hum
168	19	95.0	123	6	ABM63684	Propionib	241	19	95.0	150	3	ABG07294	Novel hum
169	19	95.0	123	6	ABM63684	Propionib	242	19	95.0	150	3	ABG07294	Novel hum
170	19	95.0	123	6	ABM63684	Propionib	243	19	95.0	150	3	ABG07294	Novel hum
171	19	95.0	123	6	ABM63684	Propionib	244	19	95.0	150	3	ABG07294	Novel hum

245	19	95.0	164	7	AD887765	Human lun	318	19	95.0	199	4	AAB689372
246	19	95.0	167	3	AAB56403	Human pro	319	19	95.0	199	4	AAB689372
247	19	95.0	167	3	ABP79188	N. gonorr	320	19	95.0	199	4	AAB31894
248	19	95.0	168	3	AAG38110	Arabidops	321	19	95.0	199	4	AAB31894
249	19	95.0	170	4	ABG26260	Novel hum	322	19	95.0	199	4	AAB31894
250	19	95.0	172	2	AAW40216	Mouse Al.	323	19	95.0	199	4	AAB31894
251	19	95.0	175	2	AAW05531	Human	324	19	95.0	199	4	AAB31894
252	19	95.0	175	2	AAW31531	Human	325	19	95.0	199	4	AAB60185
253	19	95.0	175	2	AAW20421	H. pylori	326	19	95.0	199	5	AAW47317
254	19	95.0	175	2	AAW20421	H. pylori	327	19	95.0	199	5	AAW47317
255	19	95.0	175	2	AAW20421	H. pylori	328	19	95.0	199	7	ADD46946
256	19	95.0	175	2	AAW20421	H. pylori	329	19	95.0	199	7	ADD46946
257	19	95.0	176	4	ABU89728	Human	330	19	95.0	200	4	ABG79757
258	19	95.0	176	4	ABU89728	Human	331	19	95.0	200	4	ABG79757
259	19	95.0	176	4	ABU89728	Human	332	19	95.0	200	4	ABG79757
260	19	95.0	176	4	ABU89728	Human	333	19	95.0	201	4	ABG79757
261	19	95.0	176	4	ABU89728	Human	334	19	95.0	201	4	ABG79757
262	19	95.0	176	4	ABU89728	Human	335	19	95.0	201	4	ABG79757
263	19	95.0	176	4	ABU89728	Human	336	19	95.0	201	4	ABG79757
264	19	95.0	176	4	ABU89728	Human	337	19	95.0	201	4	ABG79757
265	19	95.0	176	4	ABU89728	Human	338	19	95.0	201	4	ABG79757
266	19	95.0	176	4	ABU89728	Human	339	19	95.0	201	4	ABG79757
267	19	95.0	176	4	ABU89728	Human	340	19	95.0	201	4	ABG79757
268	19	95.0	176	4	ABU89728	Human	341	19	95.0	201	4	ABG79757
269	19	95.0	176	4	ABU89728	Human	342	19	95.0	201	4	ABG79757
270	19	95.0	176	4	ABU89728	Human	343	19	95.0	201	4	ABG79757
271	19	95.0	176	4	ABU89728	Human	344	19	95.0	201	4	ABG79757
272	19	95.0	176	4	ABU89728	Human	345	19	95.0	201	4	ABG79757
273	19	95.0	176	4	ABU89728	Human	346	19	95.0	201	4	ABG79757
274	19	95.0	176	4	ABU89728	Human	347	19	95.0	201	4	ABG79757
275	19	95.0	176	4	ABU89728	Human	348	19	95.0	201	4	ABG79757
276	19	95.0	176	4	ABU89728	Human	349	19	95.0	201	4	ABG79757
277	19	95.0	176	4	ABU89728	Human	350	19	95.0	201	4	ABG79757
278	19	95.0	176	4	ABU89728	Human	351	19	95.0	201	4	ABG79757
279	19	95.0	176	4	ABU89728	Human	352	19	95.0	201	4	ABG79757
280	19	95.0	176	4	ABU89728	Human	353	19	95.0	201	4	ABG79757
281	19	95.0	176	4	ABU89728	Human	354	19	95.0	201	4	ABG79757
282	19	95.0	176	4	ABU89728	Human	355	19	95.0	201	4	ABG79757
283	19	95.0	176	4	ABU89728	Human	356	19	95.0	201	4	ABG79757
284	19	95.0	176	4	ABU89728	Human	357	19	95.0	201	4	ABG79757
285	19	95.0	176	4	ABU89728	Human	358	19	95.0	201	4	ABG79757
286	19	95.0	176	4	ABU89728	Human	359	19	95.0	201	4	ABG79757
287	19	95.0	176	4	ABU89728	Human	360	19	95.0	201	4	ABG79757
288	19	95.0	176	4	ABU89728	Human	361	19	95.0	201	4	ABG79757
289	19	95.0	176	4	ABU89728	Human	362	19	95.0	201	4	ABG79757
290	19	95.0	176	4	ABU89728	Human	363	19	95.0	201	4	ABG79757
291	19	95.0	176	4	ABU89728	Human	364	19	95.0	201	4	ABG79757
292	19	95.0	176	4	ABU89728	Human	365	19	95.0	201	4	ABG79757
293	19	95.0	176	4	ABU89728	Human	366	19	95.0	201	4	ABG79757
294	19	95.0	176	4	ABU89728	Human	367	19	95.0	201	4	ABG79757
295	19	95.0	176	4	ABU89728	Human	368	19	95.0	201	4	ABG79757
296	19	95.0	176	4	ABU89728	Human	369	19	95.0	201	4	ABG79757
297	19	95.0	176	4	ABU89728	Human	370	19	95.0	201	4	ABG79757
298	19	95.0	176	4	ABU89728	Human	371	19	95.0	201	4	ABG79757
299	19	95.0	176	4	ABU89728	Human	372	19	95.0	201	4	ABG79757
300	19	95.0	176	4	ABU89728	Human	373	19	95.0	201	4	ABG79757
301	19	95.0	176	4	ABU89728	Human	374	19	95.0	201	4	ABG79757
302	19	95.0	176	4	ABU89728	Human	375	19	95.0	201	4	ABG79757
303	19	95.0	176	4	ABU89728	Human	376	19	95.0	201	4	ABG79757
304	19	95.0	176	4	ABU89728	Human	377	19	95.0	201	4	ABG79757
305	19	95.0	176	4	ABU89728	Human	378	19	95.0	201	4	ABG79757
306	19	95.0	176	4	ABU89728	Human	379	19	95.0	201	4	ABG79757
307	19	95.0	176	4	ABU89728	Human	380	19	95.0	201	4	ABG79757
308	19	95.0	176	4	ABU89728	Human	381	19	95.0	201	4	ABG79757
309	19	95.0	176	4	ABU89728	Human	382	19	95.0	201	4	ABG79757
310	19	95.0	176	4	ABU89728	Human	383	19	95.0	201	4	ABG79757
311	19	95.0	176	4	ABU89728	Human	384	19	95.0	201	4	ABG79757
312	19	95.0	176	4	ABU89728	Human	385	19	95.0	201	4	ABG79757
313	19	95.0	176	4	ABU89728	Human	386	19	95.0	201	4	ABG79757
314	19	95.0	176	4	ABU89728	Human	387	19	95.0	201	4	ABG79757
315	19	95.0	176	4	ABU89728	Human	388	19	95.0	201	4	ABG79757
316	19	95.0	176	4	ABU89728	Human	389	19	95.0	201	4	ABG79757
317	19	95.0	176	4	ABU89728	Human	390	19	95.0	201	4	ABG79757

391	19	95.0	224	6	ADA54376	Human pro	464	19	95.0	257	2	AAR97285	Aar97285 Human	265
392	19	95.0	224	6	ADA56787	Human sec	465	19	95.0	257	2	AAW21729	Aaw21729 Nuclear m	265
393	19	95.0	224	7	ADC74046	Human sec	466	19	95.0	257	3	AAG59110	Aag59110 Arabidops	265
394	19	95.0	224	7	ADC77843	Human sec	467	19	95.0	257	4	AAQ74447	Aaq74447 Human col	265
395	19	95.0	226	4	ADM79581	Human pro	468	19	95.0	257	4	AAQ92888	Aaq92888 C glutami	265
396	19	95.0	226	4	ABG03398	Novel hum	469	19	95.0	257	7	ADD27694	Add27694 Human adi	265
397	19	95.0	226	4	ABG09001	Novel hum	470	19	95.0	257	7	ADD26681	Add26681 Human adi	265
398	19	95.0	226	6	ADA36076	Acinetoba	471	19	95.0	257	7	ADD27489	Add27489 Human adi	265
399	19	95.0	227	5	AAEL14342	Human pro	472	19	95.0	258	3	AAAB16704	Aab16704 Bacteriop	265
400	19	95.0	228	3	AAV74793	Neisseria	473	19	95.0	258	3	AAU36296	Aau36296 Pseudomon	265
401	19	95.0	228	3	AAE47282	MSF-3792	474	19	95.0	258	6	ABU38502	Abu38502 Protein e	265
402	19	95.0	228	6	ABP78051	N. gonorr	475	19	95.0	260	4	AAW78433	Aaw78433 Human pro	265
403	19	95.0	228	6	ABP80081	N. gonorr	476	19	95.0	260	4	AAU58456	Aau58456 Propionib	265
404	19	95.0	228	6	ABG75884	Mycobacte	477	19	95.0	260	4	AAU58456	Aau58456 Propionib	265
405	19	95.0	229	4	AAAG81949	S. epider	478	19	95.0	260	6	AAU58456	Aau58456 Propionib	265
406	19	95.0	229	4	AAU04888	Micromono	479	19	95.0	261	4	AAU00701	Aau00701 Mouse FCT	265
407	19	95.0	229	5	ABG91524	Herbicida	480	19	95.0	261	5	ABG79646	Abg79646 Mouse FCT	265
408	19	95.0	230	4	AAAB80078	Corynebac	481	19	95.0	261	5	ABG78505	Abg78505 Murine FCT	265
409	19	95.0	230	4	AAAG90073	C. glutami	482	19	95.0	261	5	AAU82959	Aau82959 Human hom	265
410	19	95.0	231	2	AAW60988	Streptococ	483	19	95.0	261	6	ABU41453	Abu41453 Protein e	265
411	19	95.0	231	5	AAAB50143	Listeria	484	19	95.0	261	7	AAE54844	Aae54844 Rat Prote	265
412	19	95.0	231	6	ADA36637	Acinetoba	485	19	95.0	262	4	ABG04663	Abg04663 Novel hum	265
413	19	95.0	232	2	AAV27368	Group B S	486	19	95.0	262	5	AAU72572	Aau72572 Arabidops	265
414	19	95.0	232	3	AAAG11258	Arabidops	487	19	95.0	263	2	AAW95344	Aaw95344 Human adu	265
415	19	95.0	234	3	AAAG22116	Maize glu	488	19	95.0	263	4	ABG90550	Abg90550 Human sec	265
416	19	95.0	234	3	AAAG26676	Arabidops	489	19	95.0	263	5	ABG65464	Abg65464 Human alb	265
417	19	95.0	234	5	ABG91390	Herbicida	490	19	95.0	263	5	AAU44959	Aau44959 Human pol	265
418	19	95.0	234	5	ABG91390	Herbicida	491	19	95.0	263	5	AAU44959	Aau44959 Human pol	265
419	19	95.0	236	3	AAAB42087	Human ORF	492	19	95.0	264	3	AAAB15729	Aab15729 Arabidops	265
420	19	95.0	236	3	AAU36145	Klebsiell	493	19	95.0	264	4	AAAB86363	Aab86363 Human cer	265
421	19	95.0	236	4	AAAB61537	Degp prot	494	19	95.0	264	4	ABG18250	Abg18250 Novel hum	265
422	19	95.0	236	4	ABG27447	Novel hum	495	19	95.0	265	2	AAAR08269	Aar08269 Mouse ser	265
423	19	95.0	238	3	AAAB11695	Mouse ser	496	19	95.0	265	2	AAW70534	Aaw70534 Mouse neu	265
424	19	95.0	238	3	AAAB28132	Glycerald	497	19	95.0	265	2	AAU24490	Aau24490 Protein e	265
425	19	95.0	238	4	AAAB61538	Degp prot	498	19	95.0	266	4	AAW79417	Aaw79417 Human pro	265
426	19	95.0	238	4	AAAG83029	S. epider	499	19	95.0	267	2	AAW85606	Aaw85606 Secreted	265
427	19	95.0	238	4	AAAG83016	S. epider	500	19	95.0	267	2	AAU50254	Aau50254 Protein e	265
428	19	95.0	238	5	AAAB53557	Lactococc	501	19	95.0	268	5	AAE15256	Aae15256 Human ROL	265
429	19	95.0	238	5	AAU80750	Eucalyptu	502	19	95.0	268	5	ABP61845	Abp61845 Human pol	265
430	19	95.0	238	5	ABU44475	Protein e	503	19	95.0	269	2	AAW55786	Aaw55786 Human aqu	265
431	19	95.0	238	6	ABU42738	Protein e	504	19	95.0	269	2	AAW94322	Aaw94322 Human aqu	265
432	19	95.0	238	6	ABM72778	Staphyloc	505	19	95.0	269	4	ABG26217	Abg26217 Novel hum	265
433	19	95.0	239	3	AAV74795	Neisseria	506	19	95.0	269	6	ABP58050	Abp58050 Human aqu	265
434	19	95.0	239	3	AAV74794	Neisseria	507	19	95.0	269	7	ADB64028	Abd64028 Human pro	265
435	19	95.0	239	3	AAAG11257	Neisseria	508	19	95.0	270	4	AAAG90189	Aag90189 C glutami	265
436	19	95.0	240	3	AAAB96778	Putative	509	19	95.0	270	6	ADA05788	Ada05788 Human NOV	265
437	19	95.0	240	3	AAAG19884	Arabidops	510	19	95.0	270	7	ADC08133	Adc08133 Rice prot	265
438	19	95.0	240	3	AAAG42937	Arabidops	511	19	95.0	271	2	AAV15227	Aav15227 Human rec	265
439	19	95.0	240	7	ADE25064	Plant gro	512	19	95.0	271	3	AAE28205	Aae28205 Novel hum	265
440	19	95.0	241	5	ABP40753	Staphyloc	513	19	95.0	271	3	AAAG31287	Aag31287 Arabidops	265
441	19	95.0	241	6	ABU28310	Protein e	514	19	95.0	271	3	AAAG22050	Aag22050 Arabidops	265
442	19	95.0	243	2	AAV86089	S. pneumo	515	19	95.0	271	4	AAAB92851	Aab92851 Human mem	265
443	19	95.0	243	5	ABP38742	Staphyloc	516	19	95.0	271	4	AAAB8446	Aab8446 Human sig	265
444	19	95.0	244	5	ABP3337	Human TSA	517	19	95.0	271	4	AAAG64357	Aag64357 Human pol	265
445	19	95.0	244	6	ABU44263	Protein e	518	19	95.0	271	5	ABP90282	Abp90282 Human pol	265
446	19	95.0	245	6	ABG10688	Novel hum	519	19	95.0	271	5	ABP90282	Abp90282 Human pol	265
447	19	95.0	245	7	ADC08337	Rice prot	520	19	95.0	271	7	ADC95444	Adc95444 E. faeciu	265
448	19	95.0	247	7	ADE86966	Human pan	521	19	95.0	272	4	AAU23799	Aau23799 Novel hum	265
449	19	95.0	248	2	AAV11160	S. pneumo	522	19	95.0	272	4	AAAM94302	Aam94302 Human rep	265
450	19	95.0	249	4	AAAB61261	Human TAN	523	19	95.0	272	4	ABG16020	Abg16020 Novel hum	265
451	19	95.0	249	4	AAAB19896	Polyketid	524	19	95.0	272	6	ABU42431	Abu42431 Protein e	265
452	19	95.0	249	6	ABU11227	Human TAN	525	19	95.0	272	6	ABM71001	Abm71001 Staphyloc	265
453	19	95.0	249	6	ABU11227	Human TAN	526	19	95.0	273	3	AAAB11596	Aab11596 Mouse ser	265
454	19	95.0	250	2	AAAB33913	Sequence	527	19	95.0	273	3	AAAB86365	Aab86365 Human cer	265
455	19	95.0	250	2	AAAB33913	Sequence	528	19	95.0	273	6	ABR42553	Abra42553 Simocycli	265
456	19	95.0	250	2	AAAB60153	Nematode-	529	19	95.0	273	6	ABR42553	Abra42553 Simocycli	265
457	19	95.0	250	2	AAAB77082	Tobacco r	530	19	95.0	274	3	AAAB00202	Aab00202 Putative	265
458	19	95.0	250	3	AAAB42541	Human ORF	531	19	95.0	274	3	AAAB00211	Aab00211 Potential	265
459	19	95.0	250	3	AAU47372	Propionib	532	19	95.0	274	3	AAAB00211	Aab00211 Potential	265
460	19	95.0	250	6	ABM43891	Propionib	533	19	95.0	274	3	AAAB63322	Aab63322 Human bre	265
461	19	95.0	253	4	AAU34686	E. coli c	534	19	95.0	274	6	ABU23635	Abu23635 Protein e	265
462	19	95.0	253	7	ADC95445	E. faeciu	535	19	95.0	274	6	ABU23635	Abu23635 Protein e	265
463	19	95.0	254	6	ADA05790	Human NOV	536	19	95.0	274	7	ADC32869	Adc32869 Human nov	265

537	19	95.0	275	6	ABU24121	Protein e	610	19	95.0	300	7	ADC31643	Adc31643 Human nov
538	19	95.0	276	3	ABU42611	Human ORF	611	19	95.0	301	4	AG82151	S. epider
539	19	95.0	277	3	AAW18875	Babesia e	612	19	95.0	302	3	AG15595	Arabidops
540	19	95.0	277	3	AG22852	Arabidops	613	19	95.0	303	3	AG20322	Arabidops
541	19	95.0	278	3	AG04958	Arabidops	614	19	95.0	303	4	AG89980	C glutami
542	19	95.0	278	3	AG07719	Arabidops	615	19	95.0	304	3	AG37633	Arabidops
543	19	95.0	278	3	AG39551	Arabidops	616	19	95.0	304	3	AG04957	Arabidops
544	19	95.0	278	3	AG37634	Arabidops	617	19	95.0	304	3	AG30063	Arabidops
545	19	95.0	278	3	AG37634	Arabidops	618	19	95.0	304	3	AG07718	Arabidops
546	19	95.0	278	3	ABP70329	Streptoco	619	19	95.0	304	3	AG39550	Arabidops
547	19	95.0	278	3	ABP77659	N. gonorr	620	19	95.0	305	3	AG30713	Arabidops
548	19	95.0	278	3	ABU37537	Protein e	621	19	95.0	305	3	AG10620	Arabidops
549	19	95.0	278	3	ABU37837	Protein e	622	19	95.0	305	4	AG92837	Arabidops
550	19	95.0	279	3	AG54754	Arabidops	623	19	95.0	306	3	AG40089	Arabidops
551	19	95.0	279	3	AG54754	Arabidops	624	19	95.0	306	3	AG13663	Arabidops
552	19	95.0	279	3	ABBS091	Novel hum	625	19	95.0	306	3	AGU3620	Group B S
553	19	95.0	280	2	AAW03521	Lactococc	626	19	95.0	306	5	ABP29888	Streptoco
554	19	95.0	280	2	AAW37073	Bali modi	627	19	95.0	306	5	ABP28497	Streptoco
555	19	95.0	281	3	AG90751	C glutami	628	19	95.0	307	7	ADP70040	C. neofor
556	19	95.0	281	3	AG26675	Arabidops	629	19	95.0	307	7	ADP7261	Nuclear f
557	19	95.0	281	5	ABG77963	Ehrlichia	630	19	95.0	308	3	ABBS3390	Human col
558	19	95.0	281	6	ADA09791	E. canis	631	19	95.0	308	4	AG92974	C glutami
559	19	95.0	282	2	AAW34205	Streptomy	632	19	95.0	308	5	ABBS4256	Lactococc
560	19	95.0	282	2	AAW5806	Streptomy	633	19	95.0	308	5	ABP60904	Lactococc
561	19	95.0	282	3	AAW37116	Human pro	634	19	95.0	308	5	ABP65647	Bifidobac
562	19	95.0	282	3	AAW37116	Human pro	635	19	95.0	308	5	ABP65647	Bifidobac
563	19	95.0	282	3	AAW37116	Human pro	636	19	95.0	308	5	ABP65647	Bifidobac
564	19	95.0	282	3	AAW37116	Human pro	637	19	95.0	308	5	ABP65647	Bifidobac
565	19	95.0	282	3	AAW37116	Human pro	638	19	95.0	308	5	ABP65647	Bifidobac
566	19	95.0	282	3	AAW37116	Human pro	639	19	95.0	308	5	ABP65647	Bifidobac
567	19	95.0	282	3	AAW37116	Human pro	640	19	95.0	308	5	ABP65647	Bifidobac
568	19	95.0	282	3	AAW37116	Human pro	641	19	95.0	308	5	ABP65647	Bifidobac
569	19	95.0	282	3	AAW37116	Human pro	642	19	95.0	308	5	ABP65647	Bifidobac
570	19	95.0	282	3	AAW37116	Human pro	643	19	95.0	308	5	ABP65647	Bifidobac
571	19	95.0	282	3	AAW37116	Human pro	644	19	95.0	308	5	ABP65647	Bifidobac
572	19	95.0	282	3	AAW37116	Human pro	645	19	95.0	308	5	ABP65647	Bifidobac
573	19	95.0	282	3	AAW37116	Human pro	646	19	95.0	308	5	ABP65647	Bifidobac
574	19	95.0	282	3	AAW37116	Human pro	647	19	95.0	308	5	ABP65647	Bifidobac
575	19	95.0	282	3	AAW37116	Human pro	648	19	95.0	308	5	ABP65647	Bifidobac
576	19	95.0	282	3	AAW37116	Human pro	649	19	95.0	308	5	ABP65647	Bifidobac
577	19	95.0	282	3	AAW37116	Human pro	650	19	95.0	308	5	ABP65647	Bifidobac
578	19	95.0	282	3	AAW37116	Human pro	651	19	95.0	308	5	ABP65647	Bifidobac
579	19	95.0	282	3	AAW37116	Human pro	652	19	95.0	308	5	ABP65647	Bifidobac
580	19	95.0	282	3	AAW37116	Human pro	653	19	95.0	308	5	ABP65647	Bifidobac
581	19	95.0	282	3	AAW37116	Human pro	654	19	95.0	308	5	ABP65647	Bifidobac
582	19	95.0	282	3	AAW37116	Human pro	655	19	95.0	308	5	ABP65647	Bifidobac
583	19	95.0	282	3	AAW37116	Human pro	656	19	95.0	308	5	ABP65647	Bifidobac
584	19	95.0	282	3	AAW37116	Human pro	657	19	95.0	308	5	ABP65647	Bifidobac
585	19	95.0	282	3	AAW37116	Human pro	658	19	95.0	308	5	ABP65647	Bifidobac
586	19	95.0	282	3	AAW37116	Human pro	659	19	95.0	308	5	ABP65647	Bifidobac
587	19	95.0	282	3	AAW37116	Human pro	660	19	95.0	308	5	ABP65647	Bifidobac
588	19	95.0	282	3	AAW37116	Human pro	661	19	95.0	308	5	ABP65647	Bifidobac
589	19	95.0	282	3	AAW37116	Human pro	662	19	95.0	308	5	ABP65647	Bifidobac
590	19	95.0	282	3	AAW37116	Human pro	663	19	95.0	308	5	ABP65647	Bifidobac
591	19	95.0	282	3	AAW37116	Human pro	664	19	95.0	308	5	ABP65647	Bifidobac
592	19	95.0	282	3	AAW37116	Human pro	665	19	95.0	308	5	ABP65647	Bifidobac
593	19	95.0	282	3	AAW37116	Human pro	666	19	95.0	308	5	ABP65647	Bifidobac
594	19	95.0	282	3	AAW37116	Human pro	667	19	95.0	308	5	ABP65647	Bifidobac
595	19	95.0	282	3	AAW37116	Human pro	668	19	95.0	308	5	ABP65647	Bifidobac
596	19	95.0	282	3	AAW37116	Human pro	669	19	95.0	308	5	ABP65647	Bifidobac
597	19	95.0	282	3	AAW37116	Human pro	670	19	95.0	308	5	ABP65647	Bifidobac
598	19	95.0	282	3	AAW37116	Human pro	671	19	95.0	308	5	ABP65647	Bifidobac
599	19	95.0	282	3	AAW37116	Human pro	672	19	95.0	308	5	ABP65647	Bifidobac
600	19	95.0	282	3	AAW37116	Human pro	673	19	95.0	308	5	ABP65647	Bifidobac
601	19	95.0	282	3	AAW37116	Human pro	674	19	95.0	308	5	ABP65647	Bifidobac
602	19	95.0	282	3	AAW37116	Human pro	675	19	95.0	308	5	ABP65647	Bifidobac
603	19	95.0	282	3	AAW37116	Human pro	676	19	95.0	308	5	ABP65647	Bifidobac
604	19	95.0	282	3	AAW37116	Human pro	677	19	95.0	308	5	ABP65647	Bifidobac
605	19	95.0	282	3	AAW37116	Human pro	678	19	95.0	308	5	ABP65647	Bifidobac
606	19	95.0	282	3	AAW37116	Human pro	679	19	95.0	308	5	ABP65647	Bifidobac
607	19	95.0	282	3	AAW37116	Human pro	680	19	95.0	308	5	ABP65647	Bifidobac
608	19	95.0	282	3	AAW37116	Human pro	681	19	95.0	308	5	ABP65647	Bifidobac
609	19	95.0	282	3	AAW37116	Human pro	682	19	95.0	308	5	ABP65647	Bifidobac

683	19	95.0	312	7	ADC85589	Adc85589 Human GPC	756	19	95.0	328	4	AU35342	Au35342 Enterococ
684	19	95.0	312	7	ADD13513	Add13513 C. glutam	757	19	95.0	328	5	ABB91780	Abb91780 Herbicida
685	19	95.0	313	5	AAX37346	Aax37346 PEP PM. 1	758	19	95.0	328	6	ABO23285	AbO23285 Mouse ary
686	19	95.0	313	5	ABB49170	Abb49170 Listeria	759	19	95.0	328	6	ABO23284	AbO23284 Cow aryl-
687	19	95.0	314	3	AA929818	Aay929818 Wheat par	760	19	95.0	328	6	ABU28954	Abu28954 Protein e
688	19	95.0	314	6	ABRO1619	Abro1619 Human G p	761	19	95.0	328	6	ABM73204	Abm73204 Staphyloc
689	19	95.0	314	6	ABRO1585	Abro1585 Human G p	762	19	95.0	329	2	AAW80623	Aaw80623 S. pneumo
690	19	95.0	315	6	ABU17461	Abu17461 Protein e	763	19	95.0	329	7	ABF27377	Abf27377 Streptoco
691	19	95.0	315	4	ABBS2707	Abbs2707 Escherich	764	19	95.0	329	7	ABM73892	Abm73892 DNA clone
692	19	95.0	315	4	AAAG71585	Aag71585 Human olf	765	19	95.0	331	2	AAAG71504	Aag71504 Protein i
693	19	95.0	315	4	AAAG72285	Aag72285 Human olf	766	19	95.0	331	3	AAAG43812	Aag43812 Arabidops
694	19	95.0	315	4	AAAG72890	Aag72890 Human olf	767	19	95.0	331	3	AAU36168	Aau36168 Klebsiell
695	19	95.0	315	5	AAAE17492	Aae17492 Human sec	768	19	95.0	331	4	AAAG72652	Aag72652 Murine OR
696	19	95.0	315	5	AAAB71213	Aab71213 Human GPC	769	19	95.0	334	4	AAAS60556	Aas60556 Human gen
697	19	95.0	315	6	ABM70746	Abm70746 Photorhab	770	19	95.0	334	5	ABP41114	Abp41114 Human ova
698	19	95.0	315	6	ABRO1616	Abro1616 Human G p	771	19	95.0	334	5	ABG33878	Abg33878 Human sec
699	19	95.0	315	7	ADC37263	Adc37263 Nuclear f	772	19	95.0	334	6	ADA57046	Ada57046 Human sec
700	19	95.0	315	7	ADD12827	Add12827 Novel hum	773	19	95.0	334	6	ADA40900	Ada40900 Human sec
701	19	95.0	315	7	ADE90061	Ade90061 Human rho	774	19	95.0	334	6	ABU43221	Abu43221 Protein e
702	19	95.0	316	3	AAAG20321	Aag20321 Arabidops	775	19	95.0	334	6	ABR61839	AbR61839 Arabidops
703	19	95.0	316	3	AAU55622	Aau55622 Propionib	776	19	95.0	334	6	AAAS7736	Aas7736 Hyprsns
704	19	95.0	316	6	ABP98031	Abp98031 Protein p	777	19	95.0	335	2	AAAS7736	Aas7736 Hyprsns
705	19	95.0	316	6	ABMS2141	Abms2141 Propionib	778	19	95.0	335	3	AAAG7079	Aag7079 Human sec
706	19	95.0	317	2	AAAY01381	Aay01381 Rat osteo	779	19	95.0	336	3	AAAG32249	Aag32249 Arabidops
707	19	95.0	317	3	ABAB1662	Abab1662 Human ORF	780	19	95.0	336	3	AAAG32249	Aag32249 Arabidops
708	19	95.0	317	6	ABM69850	Abm69850 Photorhab	781	19	95.0	336	6	ABU39843	Abu39843 Protein e
709	19	95.0	317	7	ADES7454	Ades7454 Rat Prote	782	19	95.0	336	6	ABU50329	Abu50329 Protein e
710	19	95.0	318	2	AAW21746	Aaw21746 E3330-bin	783	19	95.0	337	3	AAAG7495	Aag7495 Arabidops
711	19	95.0	318	2	AAW48894	Aaw48894 Homo sapi	784	19	95.0	337	3	AAAG43811	Aag43811 Arabidops
712	19	95.0	318	2	AAW52864	Aaw52864 Apurinic/	785	19	95.0	337	3	AAAG6421	Aag6421 Arabidops
713	19	95.0	318	3	AAAG30062	Aag30062 Arabidops	786	19	95.0	337	3	AAAG40549	Aag40549 Arabidops
714	19	95.0	318	3	AAAG08737	Aag08737 Arabidops	787	19	95.0	337	3	ADA35864	Ada35864 Acinetoba
715	19	95.0	318	3	AAAG4921	Aag4921 Arabidops	788	19	95.0	338	2	AAAY07063	Aay07063 Renal can
716	19	95.0	318	5	ABP60935	Abp60935 Zymomonas	789	19	95.0	339	4	AAAB61273	Aab61273 Human TAN
717	19	95.0	318	7	ADD08904	Add08904 Apurinic/	790	19	95.0	339	4	AAAB61273	Aab61273 Human TAN
718	19	95.0	318	7	ADES25727	Ade25727 Human pro	791	19	95.0	339	4	AAAB61275	Aab61275 Human TAN
719	19	95.0	319	3	AAAG31049	Aag31049 Arabidops	792	19	95.0	339	4	AAAB61275	Aab61275 Human TAN
720	19	95.0	319	4	AAAB61257	Aab61257 Mature hu	793	19	95.0	339	4	AAAB61275	Aab61275 Human TAN
721	19	95.0	319	4	AAAB93403	Aab93403 Human gly	794	19	95.0	339	4	AAAY72790	Aay72790 Human pla
722	19	95.0	319	6	ABP98028	Abp98028 Protein p	795	19	95.0	339	4	AAAY72790	Aay72790 Human pla
723	19	95.0	319	6	ABU11223	Abu11223 Human TAN	796	19	95.0	339	4	AAAB38760	Aab38760 Human pol
724	19	95.0	320	3	AAAG50998	Aag50998 Arabidops	797	19	95.0	339	4	AAAB31668	Aab31668 Amiro aci
725	19	95.0	320	3	AAAG62886	Aag62886 Arabidops	798	19	95.0	339	4	AAAB31668	Aab31668 Amiro aci
726	19	95.0	320	3	AAAG40088	Aag40088 Arabidops	799	19	95.0	339	6	ABP72608	Abp72608 Benign pr
727	19	95.0	320	3	AAAG13662	Aag13662 Arabidops	800	19	95.0	339	6	ABP72607	Abp72607 Benign pr
728	19	95.0	320	3	AAAG28247	Aag28247 S. epider	801	19	95.0	339	6	ADA84105	Ada84105 Human GPC
729	19	95.0	320	5	ABBS1676	Abbs1676 Herbicida	802	19	95.0	339	6	ABU11221	Abu11221 Human TAN
730	19	95.0	321	3	AAAY99817	Aay99817 Soybean p	803	19	95.0	339	6	ABU11221	Abu11221 Human TAN
731	19	95.0	321	5	ABP51575	Abp51575 Human G p	804	19	95.0	339	6	ABU11240	Abu11240 Glycoprot
732	19	95.0	321	6	ADA89479	Ada89479 Staphyloc	805	19	95.0	339	6	ABU11241	Abu11241 Glycoprot
733	19	95.0	322	3	AAAY94668	Aay94668 Murine unc	806	19	95.0	339	6	ABU11242	Abu11242 Glycoprot
734	19	95.0	322	3	AAAY94668	Aay94668 Murine unc	807	19	95.0	339	6	ABU70998	Abu70998 Human adi
735	19	95.0	322	3	AAAG20529	Aag20529 Arabidops	808	19	95.0	339	6	ABR82451	AbR82451 Human ARP
736	19	95.0	322	3	AAAG08879	Aag08879 Arabidops	809	19	95.0	340	4	ABG05957	Abg05957 Pseudomon
737	19	95.0	322	3	ABU25221	Abu25221 Protein e	810	19	95.0	340	4	ABR42679	AbR42679 Escherich
738	19	95.0	322	6	ABM72033	Abm72033 Staphyloc	811	19	95.0	342	4	ABBS2472	Abbs2472 Escherich
739	19	95.0	323	3	AAAG06285	Aag06285 Arabidops	812	19	95.0	342	4	ABBS2472	Abbs2472 Escherich
740	19	95.0	323	3	ABU41982	Abu41982 Protein e	813	19	95.0	342	5	ABG79682	Abg79682 Tumour in
741	19	95.0	324	3	AAAG06284	Aag06284 Arabidops	814	19	95.0	343	3	AAAG15727	Aag15727 Arabidops
742	19	95.0	324	6	ABU33384	Abu33384 Protein e	815	19	95.0	343	3	AAAG45273	Aag45273 Arabidops
743	19	95.0	324	6	ABU29258	Abu29258 Protein e	816	19	95.0	343	5	ABBS1220	Abbs1220 Herbicida
744	19	95.0	324	6	ADA33349	Ada33349 Acinetoba	817	19	95.0	344	3	AAAG22851	Aag22851 Arabidops
745	19	95.0	325	2	AAAY50034	Aay50034 Porcine c	818	19	95.0	344	3	AAAG22851	Aag22851 Arabidops
746	19	95.0	325	3	AAAY94665	Aay94665 Murine unc	819	19	95.0	345	7	ADC94222	Adc94222 E. faeciu
747	19	95.0	325	3	AAAY94665	Aay94665 Human unc	820	19	95.0	346	3	AAAB32505	Aab32505 S. lavend
748	19	95.0	325	3	AAAG20320	Aag20320 Arabidops	821	19	95.0	346	7	ADE10277	Ade10277 S. lavend
749	19	95.0	326	6	ADA05792	Ada05792 Human NOV	822	19	95.0	347	5	AAAG3168	Aag3168 C glutami
750	19	95.0	326	3	AAAB29339	Aab29339 Hepatitis	823	19	95.0	347	5	ABP38453	Abp38453 Staphyloc
751	19	95.0	326	3	AAAG46264	Aag46264 Arabidops	824	19	95.0	347	5	ABP38453	Abp38453 Staphyloc
752	19	95.0	326	3	AAAG51053	Aag51053 Arabidops	825	19	95.0	347	5	ABP41750	Abp41750 Human ova
753	19	95.0	326	5	ABBS9190	Abbs9190 Herbicida	826	19	95.0	347	5	ADC54077	Adc54077 Mouse tes
754	19	95.0	327	2	AAV50033	Aay50033 Porcine c	827	19	95.0	347	7	ADC54079	Adc54079 Mouse tes
755	19	95.0	327	6	ABU11347	Abu11347 Protein e	828	19	95.0	347	7	ADC54079	Adc54079 Mouse tes

829	19	95.0	350	3	AA02086	Abu02086 F420-depe	902	19	95.0	377	6	ABU25867	Abu25867 Protein e
830	19	95.0	350	5	AA015202	Aac15202 Rhodococc	903	19	95.0	378	4	AA093361	Aam93361 Human pol
831	19	95.0	350	5	AAU97147	Aau97147 Rhodococc	904	19	95.0	378	5	AA020087	Aae20087 Lactobaci
832	19	95.0	350	5	AAE16748	Aae16748 Rhodococc	905	19	95.0	379	6	ABU23947	Abu23947 Protein e
833	19	95.0	351	6	ABU18964	Abu18964 Protein e	906	19	95.0	379	6	ABU49216	Abu49216 Protein e
834	19	95.0	352	6	ABU48732	Abu48732 Listeria	907	19	95.0	379	7	ABU04817	Abu04817 Rat cdc 3
835	19	95.0	352	6	ABU32749	Abu32749 Protein e	908	19	95.0	379	7	AD445256	Ad445256 Rat cdc 3
836	19	95.0	353	3	AAU94667	Aau94667 Human unc	909	19	95.0	380	2	AAU98800	Aau98800 H. pylori
837	19	95.0	353	3	AAU08899	Aau08899 Arabidops	910	19	95.0	380	4	AAU35898	Aau35898 Helicobac
838	19	95.0	353	4	ABG16162	Abg16162 Novel hum	911	19	95.0	380	5	ABG30683	Abg30683 Cystathio
839	19	95.0	354	3	AAU3810	Aau3810 Arabidops	912	19	95.0	380	5	ABU33267	Abu33267 Protein e
840	19	95.0	355	4	AAU34801	Aau34801 Arabidops	913	19	95.0	382	5	ABP43943	Abp43943 Unidentif
841	19	95.0	355	4	AAU34802	Aau34802 Human col	914	19	95.0	382	6	ABU33959	Abu33959 Protein e
842	19	95.0	356	6	AAU33382	Aau33382 HTRM clon	915	19	95.0	383	4	ABU33959	Abu33959 Protein e
843	19	95.0	356	6	AD36384	Ada36384 Acinetoba	916	19	95.0	383	6	ABU60853	Abu60853 Mevalonat
844	19	95.0	357	3	AAU29807	Aau29807 Arabidops	917	19	95.0	383	6	ABU15985	Abu15985 Protein e
845	19	95.0	357	4	AAU29807	Aau29807 Arabidops	918	19	95.0	383	6	ABU15985	Abu15985 Protein e
846	19	95.0	357	5	ABP63001	Abp63001 Human pol	919	19	95.0	383	6	ABU15985	Abu15985 Protein e
847	19	95.0	357	5	ABP63001	Abp63001 Human pol	920	19	95.0	384	3	AAU71056	Aau71056 Human mem
848	19	95.0	358	3	AAU29806	Aau29806 Arabidops	921	19	95.0	384	4	AAU40198	Aau40198 Human pol
849	19	95.0	359	3	AAU15679	Aau15679 Arabidops	922	19	95.0	384	4	AAU40198	Aau40198 Human pol
850	19	95.0	359	3	AAU32423	Aau32423 Arabidops	923	19	95.0	384	5	ABU94802	Abu94802 Human pro
851	19	95.0	359	3	AAU48953	Aau48953 Arabidops	924	19	95.0	384	5	ABU94802	Abu94802 Human pro
852	19	95.0	360	4	AAU40979	Aau40979 Human pol	925	19	95.0	384	6	ABU94802	Abu94802 Human pro
853	19	95.0	360	7	AAU09991	Aau09991 Novel pro	926	19	95.0	385	5	AAU72542	Aau72542 Arabidops
854	19	95.0	361	3	AAU08898	Aau08898 Arabidops	927	19	95.0	386	3	ABU13568	Abu13568 Streptomy
855	19	95.0	361	3	AAU45156	Aau45156 Arabidops	928	19	95.0	386	4	AAU35448	Aau35448 Haemophil
856	19	95.0	361	3	AAU08897	Aau08897 Arabidops	929	19	95.0	386	4	AAU35448	Aau35448 Haemophil
857	19	95.0	367	4	AAU13149	Aau13149 Human pol	930	19	95.0	388	5	ABG30682	Abg30682 Cystathio
858	19	95.0	367	4	AAU13149	Aau13149 Human pol	931	19	95.0	388	5	ABG30682	Abg30682 Cystathio
859	19	95.0	369	4	AAU72791	Aau72791 Human GPV	932	19	95.0	389	6	AAU99731	Aau99731 Streptomy
860	19	95.0	369	4	AAU72791	Aau72791 Human GPV	933	19	95.0	389	6	AAU99731	Aau99731 Streptomy
861	19	95.0	369	4	AAU72791	Aau72791 Human GPV	934	19	95.0	389	6	AAU99731	Aau99731 Streptomy
862	19	95.0	369	4	AAU72791	Aau72791 Human GPV	935	19	95.0	389	6	AAU99731	Aau99731 Streptomy
863	19	95.0	369	5	AAU28223	Aau28223 Streptoco	936	19	95.0	390	6	ABU10407	Abu10407 Arthrobac
864	19	95.0	369	6	AAU01488	Aau01488 S. pneumo	937	19	95.0	390	7	ABU10407	Abu10407 Arthrobac
865	19	95.0	369	6	AAU01488	Aau01488 S. pneumo	938	19	95.0	391	7	ABU10407	Abu10407 Arthrobac
866	19	95.0	369	6	AAU01488	Aau01488 S. pneumo	939	19	95.0	391	7	ABU10407	Abu10407 Arthrobac
867	19	95.0	369	6	AAU01488	Aau01488 S. pneumo	940	19	95.0	391	7	ABU10407	Abu10407 Arthrobac
868	19	95.0	369	7	AAU01488	Aau01488 S. pneumo	941	19	95.0	391	7	ABU10407	Abu10407 Arthrobac
869	19	95.0	369	7	AAU01488	Aau01488 S. pneumo	942	19	95.0	391	7	ABU10407	Abu10407 Arthrobac
870	19	95.0	370	4	AAU00700	Aau00700 Mouse FCT	943	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
871	19	95.0	370	4	AAU00700	Aau00700 Mouse FCT	944	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
872	19	95.0	370	4	AAU00700	Aau00700 Mouse FCT	945	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
873	19	95.0	370	5	AAU00700	Aau00700 Mouse FCT	946	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
874	19	95.0	370	5	AAU00700	Aau00700 Mouse FCT	947	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
875	19	95.0	370	5	AAU00700	Aau00700 Mouse FCT	948	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
876	19	95.0	370	5	AAU00700	Aau00700 Mouse FCT	949	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
877	19	95.0	370	5	AAU00700	Aau00700 Mouse FCT	950	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
878	19	95.0	370	6	AAU00700	Aau00700 Mouse FCT	951	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
879	19	95.0	370	6	AAU00700	Aau00700 Mouse FCT	952	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
880	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	953	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
881	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	954	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
882	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	955	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
883	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	956	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
884	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	957	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
885	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	958	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
886	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	959	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
887	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	960	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
888	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	961	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
889	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	962	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
890	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	963	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
891	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	964	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
892	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	965	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
893	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	966	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
894	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	967	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
895	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	968	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
896	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	969	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
897	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	970	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
898	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	971	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
899	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	972	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
900	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	973	19	95.0	392	6	ABU10407	Abu10407 Arthrobac
901	19	95.0	371	5	AAU00700	Aau00700 Mouse FCT	974	19	95.0	392	6	ABU10407	Abu10407 Arthrobac

975 19 95.0 403 3 AAG50762 Arabidops  
976 19 95.0 403 3 AAG13524 Arabidops  
977 19 95.0 403 4 ABB58218 Drosophil  
978 19 95.0 404 3 AAG53518 Arabidops  
979 19 95.0 404 3 AAG39960 Arabidops  
980 19 95.0 404 6 ABU40766 Protein e  
981 19 95.0 405 2 AAY14954 Amino aci  
982 19 95.0 405 4 AAB83254 Rat FATP4  
983 19 95.0 405 6 ABM69291 Phototrab  
984 19 95.0 405 4 AAR21549 Human Try  
985 19 95.0 406 2 ABM23329 Protein e  
986 19 95.0 407 4 AAM33361 Human pol  
987 19 95.0 407 5 ABP27964 Streptoco  
988 19 95.0 407 5 ABP61471 Human NF-  
989 19 95.0 407 6 ABU46908 Protein e  
990 19 95.0 408 4 AAG91382 C glutami  
991 19 95.0 409 5 AAE30494 Fruit fly  
992 19 95.0 410 2 AAR54663 Transcrip  
993 19 95.0 410 2 AAR89212 Transcrip  
994 19 95.0 410 5 AAE30503 Fruit fly  
995 19 95.0 411 4 AAU56508 Propionib  
996 19 95.0 411 6 ABM53027 Propionib  
997 19 95.0 411 6 ABU50525 Protein e  
998 19 95.0 412 3 AAB58972 Breast an  
999 19 95.0 412 3 AAB57091 Human pro  
1000 19 95.0 412 6 ABU14941 Protein e

ALIGNMENTS

3SULT 1  
AAW47151  
J AAM47151 standard; peptide; 6 AA.  
K  
L AAM47151;  
M  
N 12-FEB-2002 (first entry)  
O  
P S chrysomallus actinomycin biosynthesis protein acmC fragment #9.  
Q Modular enzyme system; cyclic gene synthesis; repetitive coding sequence;  
R antibiotic; non-ribosomal peptide synthetase; NRPS; PKS;  
S polyketide synthase; actinomycin biosynthesis.  
T Streptomyces chrysomallus.  
U Synthetic.  
V WO200181564-A2.  
W  
X 01-NOV-2001.  
Y 25-APR-2001; 2001WO-DE001578.  
Z  
26-APR-2000; 2000DE-01021267.  
(ACTI-) ACTINODRUG PHARM GMBH.  
Schauwecker F;  
WPI; 2002-049276/06.  
N-PSDB; ABA03345.  
Preparing DNA encoding modular protein for e.g. producing new enzymes for  
synthesis of polyketide antibiotics, comprises cyclic integration of  
fragments into a vector.  
Example 3; Page 54; 83pp; German.  
The present invention relates to the preparation of DNA, in a circular  
vector, that encodes one or more segments of a modular polypeptide. DNA  
or DNA libraries produced this way are used to produce modular  
polypeptides, particularly enzymes, which can be used to act on

CC substrates to produce compounds for therapeutic testing. Enzymes of  
CC particular interest are those involved in non-ribosomal peptide synthesis  
CC or polyketide synthesis, and compounds for testing are particularly  
CC macrolide antibiotics, including penicillins, vancomycins or  
CC erythromycins, but may also be modular receptors. The present sequence is  
CC a fragment of a protein encoded by a Streptomyces chrysomallus  
CC actinomycin biosynthesis gene which was used in a plasmid in the  
CC exemplification of the invention  
SQ Sequence 6 AA;  
Query Match 95.0%; Score 19; DB 5; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 VAEF 5  
Db 3 VAEF 6  
RESULT 2  
AAB07871  
ID AAB07871 standard; peptide; 7 AA.  
XX  
AC AAB07871;  
XX 14-NOV-2000 (first entry)  
XX  
DE A beta-secretase inhibitor peptide.  
XX  
KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
XX  
OS Synthetic.  
FH Key Location/Qualifiers  
FT Modified-site 3 /note= "hydroxyethylene"  
FT  
XX WO200047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PF 10-FEB-2000; 2000WO-US003819.  
XX  
PR 10-FEB-1999; 99US-0119571P.  
PR 15-JUN-1999; 99US-0139172P.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Sinha S, Tataeno G, Tung J, Wang S, McConlogue L;  
XX  
DR WPI; 2000-533011/48.  
XX  
PT Purified beta-secretase protein used in assays to discover inhibitors  
PT which can be used for the treatment of amyloidogenic diseases e.g.  
PT Alzheimer's disease.  
XX  
PS Disclosure; Page 12; 121pp; English.  
XX  
CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
CC enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to a  
CC mammalian subject e.g. with Alzheimer's disease or Alzheimer's disease-  
CC like pathology to test if they maintain or improve cognitive ability or  
CC reduce the plaque burden. The compounds are used for the treatment of  
CC amyloidogenic diseases e.g. Alzheimer's disease. The present sequence  
CC represents an inhibitor of beta-secretase enzyme  
XX

SQ Sequence 7 AA;

Query Match 95.0%; Score 19; DB 3; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 2 VAEF 5  
 ||||  
 Db 4 VAEF 7

RESULT 3

AA07872  
 ID AAB07872 standard; peptide; 8 AA.

XX

AC

XX

DT

XX

XX

DE

XX

KW

KW

KW

OS

XX

XX

FH

FT

FT

XX

XX

PN

XX

PD

XX

XX

PF

PR

PR

PR

XX

XX

PA

XX

XX

PI

PI

PI

XX

XX

DR

XX

XX

PT

PT

PT

XX

XX

PS

XX

XX

CC

CC

CC

CC

CC

CC

CC

XX

SQ Sequence 7 AA;

Query Match 95.0%; Score 19; DB 3; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 2 VAEF 5  
 ||||  
 Db 4 VAEF 7

RESULT 3

AA07872  
 ID AAB07872 standard; peptide; 8 AA.

XX

AC

XX

DT

XX

XX

DE

XX

KW

KW

KW

OS

XX

XX

FH

FT

FT

XX

XX

PN

XX

PD

XX

XX

PF

PR

PR

PR

XX

XX

PA

XX

XX

PI

PI

PI

XX

XX

DR

XX

XX

PT

PT

PT

XX

XX

PS

XX

XX

CC

CC

CC

CC

CC

CC

CC

XX

SQ Sequence 7 AA;

Query Match 95.0%; Score 19; DB 3; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 2 VAEF 5  
 ||||  
 Db 4 VAEF 7

RESULT 3

AA07872  
 ID AAB07872 standard; peptide; 8 AA.

XX

AC

XX

DT

XX

XX

DE

XX

KW

KW

KW

OS

XX

XX

FH

FT

FT

XX

XX

PN

XX

PD

XX

XX

PF

PR

PR

PR

XX

XX

PA

XX

XX

PI

PI

PI

XX

XX

DR

XX

XX

PT

PT

PT

XX

XX

PS

XX

XX

CC

CC

CC

CC

CC

CC

CC

XX

SQ Sequence 7 AA;

Query Match 95.0%; Score 19; DB 3; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 2 VAEF 5  
 ||||  
 Db 4 VAEF 7

RESULT 3

AA07872  
 ID AAB07872 standard; peptide; 8 AA.

XX

AC

XX

DT

XX

XX

DE

XX

KW

KW

KW

OS

XX

XX

FH

FT

FT

XX

XX

PN

XX

PD

XX

XX

PF

PR

PR

PR

XX

XX

PA

XX

XX

PI

PI

PI

XX

XX

DR

XX

XX

PT

PT

PT

XX

XX

PS

XX

XX

CC

CC

CC

CC

CC

CC

CC

XX

SQ Sequence 7 AA;

Query Match 95.0%; Score 19; DB 3; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 2 VAEF 5  
 ||||  
 Db 4 VAEF 7

RESULT 3

AA07872  
 ID AAB07872 standard; peptide; 8 AA.

XX

AC

XX

DT

XX

XX

DE

XX

KW

KW

KW

OS

XX

XX

FH

FT

FT

XX

XX

PN

XX

PD

XX

XX

PF

PR

PR

PR

XX

XX

PA

XX

XX

PI

PI

PI

XX

XX

DR

XX

XX

PT

PT

PT

XX

XX

PS

XX

XX

CC

CC

CC

CC

CC

CC

CC

XX

SQ Sequence 7 AA;

Query Match 95.0%; Score 19; DB 3; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 2 VAEF 5  
 ||||  
 Db 4 VAEF 7

RESULT 3

AA07872  
 ID AAB07872 standard; peptide; 8 AA.

XX

AC

XX

DT

XX

XX

DE

XX

KW

KW

KW

OS

XX

XX

FH

FT

FT

XX

XX

PN

XX

PD

XX

XX

PF

PR

PR

PR

XX

XX

PA

XX

XX

PI

PI

PI

XX

XX

DR

XX

XX

PT

PT

PT

XX

XX

PS

XX

XX

CC

XX DPI tryptic digest peptide #317.

DE Human; depression associated protein isoform; tryptic digest peptide;  
 KW DPI; cerebrospinal fluid; CSF; BAD; bipolar affective disorder;  
 KW neuropsychiatric disorder; bipolar mood disorder; neuroleptic;  
 KW maniac-depressive illness; schizoaffective disorder.  
 XX Homo sapiens.  
 OS WO200162785-A2.  
 PN 30-AUG-2001.  
 XX 23-FEB-2001; 2001WO-GB000786.  
 XX 24-FEB-2000; 2000GB-00004412.  
 PR 08-DEC-2000; 2000GB-00030050.  
 PR 12-DEC-2000; 2000US-0254830P.  
 XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
 PA Herath HMAC, Parekh RB, Rohlf C, Terrett JA, Tyson KL;  
 PI WPI; 2001-570626/64.  
 XX Novel nucleic acid encoding a protein associated with bipolar affective  
 PT disorder, which is used for diagnosis, prophylaxis and therapy of  
 PT neuropsychiatric disorders, such as bipolar affective disorder.  
 XX Disclosure; Page 37; 153pp; English.

XX The present invention relates to the identification of depression  
 CC associated protein isoforms (DPIs), particularly the tryptic digest  
 CC peptides of these proteins. Some of the DPIs (AAU28404-AAU28625)  
 CC described are decreased in the cerebrospinal fluid (CSF) of BAD (bipolar  
 CC affective disorder) subjects, whilst other DPIs (AAU28626-AAU28987) are  
 CC increased in BAD subjects. Also described are peptide sequences  
 CC identified from DPI-45 and DPI-213 and the nucleic acid sequence they are  
 CC encoded by. The sequences of the invention are useful for clinical  
 CC screening, diagnosis, prognosis, therapy and prophylaxis of  
 CC neuropsychiatric disorders e.g. BAD (also known as bipolar mood disorder,  
 CC BP), maniac-depressive illnesses, attention deficit disorders,  
 CC schizoaffective disorders, and unipolar affective disorders. The present  
 CC sequence represents one of the DPI tryptic digest peptides of the present  
 CC invention

XX Sequence 8 AA;  
 SQ Query Match 95.0%; Score 19; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06; Mismatches 0; Gaps 0;  
 Matches 4; Conservative 0; Indels 0

QY 2 VAEF 5  
 ||||  
 Db 2 VAEF 5

RESULT 6  
 AAU25114  
 ID AAU25114 standard; peptide; 8 AA.  
 AC AAU25114;  
 XX 18-DEC-2001 (first entry)  
 DT Schizophrenia-Associated Protein Isoform (SPI) peptide #343.  
 DE Schizophrenia-associated protein isoform; SPI; SPI-206; SPI-238;  
 XX neuroleptic; gene therapy; cerebrospinal fluid; serum; plasma.  
 KW Homo sapiens.  
 OS

PN WO200162785-A2.  
 XX 30-AUG-2001.  
 XX 23-FEB-2001; 2001WO-GB000792.  
 XX 24-FEB-2000; 2000GB-00004415.  
 PR 28-DEC-2000; 2000US-00750395.  
 XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
 PA Herath HMAC, Parekh RB, Rohlf C, Terrett JA, Tyson KL;  
 PI WPI; 2001-570624/64.  
 XX New schizophrenia associated protein isoforms and encoding nucleic acid  
 PT molecules, useful for treatment, diagnosis and prognosis of schizophrenia  
 PT and screening for potential drugs for treatment and new drug targets.  
 XX Disclosure; Page 36; 149pp; English.

XX The sequence represents a schizophrenia-associated protein isoform (SPI).  
 CC These protein isoforms, e.g. SPI-206, SPI-238 and SPI-240 are detectable  
 CC in cerebrospinal fluid, serum or plasma and are useful markers of  
 CC schizophrenia. The sequences can be used for treatment and diagnosis of  
 CC schizophrenia, screening, prognosis, monitoring the results of therapy,  
 CC identifying patients most likely to respond to a particular therapy and  
 CC identification of new targets for drug treatment. SPI DNA is useful as a  
 CC nucleic acid probe to detect the presence of nucleic acids or SPIs

XX Sequence 8 AA;  
 SQ Query Match 95.0%; Score 19; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06; Mismatches 0; Gaps 0;  
 Matches 4; Conservative 0; Indels 0

QY 2 VAEF 5  
 ||||  
 Db 2 VAEF 5

RESULT 7  
 AAU26368  
 ID AAU26368 standard; peptide; 8 AA.  
 AC AAU26368;  
 XX 18-DEC-2001 (first entry)  
 DT Depression-Associated Protein isoform DPI-124.  
 DE Human; Bipolar Affective Disorder; BAD; Depression-Associated feature;  
 XX DP; Depression-Associated protein isoform; DPI; Cerebro-spinal fluid;  
 KW CSF; antidepressant; antimanic; nootropic; tranquiliser; neuroleptic;  
 KW attention deficient disorder; schizoaffective disorder;  
 KW unipolar affective disorder.  
 XX Homo sapiens.  
 OS WO200163294-A2.  
 PN 30-AUG-2001.  
 XX 23-FEB-2001; 2001WO-GB000791.  
 XX 24-FEB-2000; 2000GB-00004412.  
 PR 08-DEC-2000; 2000GB-00030050.  
 PR 12-DEC-2000; 2000US-0254830P.  
 XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
 PA Herath HMAC, Parekh RB, Rohlf C;  
 PI



CC human MAGE antigen (Ag) in a mammal comprises, a MAGE CTL Ag response  
 CC inducing peptide (i.e. AAR78904 to AAR78917) and a lipid conjugated  
 CC helper T cell inducing peptide. The compen. is useful in the treatment  
 CC and prevention of MAGE tumour Ag associated diseases, e.g. melanoma  
 CC cancers. (Updated on 25-MAR-2003 to correct FI field.)  
 XX

SQ Sequence 9 AA;

Query Match 95.0%; Score 19; DB 2; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 Db 2 VAEF 5

RESULT 10

ID ABB77871 standard; peptide; 9 AA.

XX AC ABB77871;

DT 27-SEP-2002 (first entry)

DE A beta-amyloid precursor protein (APP) inhibitor peptide.

XX Beta-amyloid precursor protein; APP; APP inhibitor peptide; BACE;  
 KW beta site APP cleaving enzyme; protein coordinate data; APP751;  
 KW Swedish family mutation; Alzheimer's disease.

XX Synthetic.

OS Homo sapiens.

XX Key Location/Qualifiers  
 FH Misc-difference 5  
 FT /label= Sta  
 FT /note= "statine"

PN W0200225276-A1.

XX 28-MAR-2002.

XX 19-SEP-2001; 2001WO-US029387.

XX 22-SEP-2000; 2000US-0234576P.

XX (AMHP ) AMERICAN HOME PROD CORP.

XX Choppa R, Svenson K, Annis B, Akopian TN, Bard J, Stahl ML;  
 PI Somers WS;

XX WPI; 2002-519081/55.

XX Crystallized complex of beta-site amyloid precursor protein (APP)  
 FT cleaving enzyme (BACE) and APP inhibitor peptide useful for identifying  
 PT agents that interact with active site of BACE or active site of APP  
 PT binding protein or peptide.

XX Claim 1; Page 22; 88pp; English.

XX The present sequence represents a beta-amyloid precursor protein (APP)  
 CC inhibitor peptide. This peptide inhibits binding between APP and BACE  
 CC (beta site APP cleaving enzyme). The specification describes a  
 CC crystallized complex of BACE and the present APP inhibitor. Protein  
 CC coordinate data for BACE is given in the specification. The APP inhibitor  
 CC peptide is based on the P10 to P4' APP751 Swedish family mutation. The  
 CC crystallized complex is used for identifying an agent that interacts with  
 CC an active site of BACE or an active site of an APP binding protein or  
 CC peptide. The agents are useful in the treatment and/or prevention of  
 CC Alzheimer's disease

XX Sequence 9 AA;

Query Match 95.0%; Score 19; DB 5; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 Db 6 VAEF 9

RESULT 11

ID ABB06593 standard; peptide; 10 AA.

XX AC ABB06593;

DT 31-MAY-2002 (first entry)

DE Beta-secretase related peptide SEQ ID NO:197.

XX Beta-secretase; enzyme; cleavage site; amyloid protein precursor; APP;  
 KW aspartyl protease; neuroprotective; nontropic; beta-secretase inhibitor;  
 KW Alzheimer's disease.

XX Homo sapiens.

OS Synthetic.

XX W0200206306-A2.

XX 24-JAN-2002.

XX 19-JUL-2001; 2001WO-US023035.

XX 19-JUL-2000; 2000US-0219795P.

XX 12-MAR-2001; 2001US-0275251P.

XX (PHAA ) PHARMACIA & UPJOHN CO.

XX Yan R, Tomasselli AG, Gurney ME, Emmons TL, Bienkowski MJ;  
 PI Heinrikson RL;

XX WPI; 2002-216995/27.

XX Novel substrates for human aspartyl protease useful for identifying  
 PT modulators of beta secretase activity of aspartyl protease for treating  
 PT Alzheimer's disease.

XX Disclosure; Page 188; 188pp; English.

XX The present invention describes an isolated peptide (I) comprising a  
 CC sequence of at least four amino acids, where the peptide is a substrate  
 CC for conducting aspartyl protease assays. (I) has neuroprotective and  
 CC nontropic activities, and can be used as an inhibitor of beta-secretase  
 CC activity. A beta-secretase modulator from the present invention can be  
 CC used for inhibiting beta-secretase activity in vivo, and in the  
 CC manufacture of a medicament for the treatment of Alzheimer's disease.  
 CC Pharmaceutical compositions from the present invention can be used for  
 CC treating a disease or condition characterised by an abnormal beta-  
 CC secretase activity. (I) is useful for identifying agents that modulate  
 CC the activity of human Asp2 aspartyl protease (Hu-Asp2). (I) is useful as  
 CC a core structure to construct derivatives. ABL4914 to ABL4925 and  
 CC ABB06409 to ABB06593 represent sequences used in the exemplification of  
 CC the present invention

XX Sequence 10 AA;

Query Match 95.0%; Score 19; DB 5; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 Db 5 VAEF 8

RESULT 12  
 ABB06592  
 ID ABB06592 standard; peptide; 12 AA.  
 AC ABB06592;  
 DT 31-MAY-2002 (first entry)  
 XX Beta-secretase related peptide SEQ ID NO:196.  
 DE Beta-secretase; enzyme; cleavage site; amyloid protein precursor; APP;  
 KW aspartyl protease; neuroprotective; nontropic; beta-secretase inhibitor;  
 KW Alzheimer's disease.  
 XX Homo sapiens.  
 JS Synthetic.  
 JS WO200206306-A2.  
 PN 24-JAN-2002.  
 PD 19-JUL-2001; 2001WO-US023035.  
 PF 19-JUL-2000; 2000US-0219795P.  
 PX 12-MAR-2001; 2001US-0275251P.  
 PR (PHAA) PHARMACIA & UPJOHN CO.  
 PX Yan R, Tomasselli AG, Gurney ME, Emmons TL, Bienkowski MJ;  
 PI Heintz RL;  
 XX WPI; 2002-216995/27.  
 XX Novel substrates for human aspartyl protease useful for identifying  
 PT modulators of beta secretase activity of aspartyl protease for treating  
 PT Alzheimer's disease.  
 XX Disclosure; Page 188; 188pp; English.  
 XX The present invention describes an isolated peptide (I) comprising a  
 CC sequence of at least four amino acids, where the peptide is a substrate  
 CC for conducting aspartyl protease assays. (I) has neuroprotective and  
 CC nontropic activities, and can be used as an inhibitor of beta-secretase  
 CC activity. A beta-secretase modulator from the present invention can be  
 CC used for inhibiting beta-secretase activity in vivo, and in the  
 CC manufacture of a medicament for the treatment of Alzheimer's disease.  
 CC Pharmaceutical compositions from the present invention can be used for  
 CC treating a disease or condition characterised by an abnormal beta-  
 CC secretase activity. (I) is useful for identifying agents that modulate  
 CC the activity of human Asp2 aspartyl protease (Hu-Asp2). (I) is useful as  
 CC a core structure to construct derivatives. ABL49914 to ABL49925 and  
 CC ABB06409 to ABB06593 represent sequences used in the exemplification of  
 CC the present invention  
 XX Sequence 12 AA;  
 SQ Query Match 95.0%; Score 19; DB 5; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 DB 5 VAEF 8  
 RESULT 13  
 ABB06592  
 ID ABB06592 standard; peptide; 12 AA.  
 AC ABB06592;  
 DT 31-MAY-2002 (first entry)  
 XX Beta-secretase related peptide SEQ ID NO:196.  
 DE Beta-secretase; enzyme; cleavage site; amyloid protein precursor; APP;  
 KW aspartyl protease; neuroprotective; nontropic; beta-secretase inhibitor;  
 KW Alzheimer's disease.  
 XX Homo sapiens.  
 JS Synthetic.  
 JS WO200206306-A2.  
 PN 24-JAN-2002.  
 PD 19-JUL-2001; 2001WO-US023035.  
 PF 19-JUL-2000; 2000US-0219795P.  
 PX 12-MAR-2001; 2001US-0275251P.  
 PR (PHAA) PHARMACIA & UPJOHN CO.  
 PX Yan R, Tomasselli AG, Gurney ME, Emmons TL, Bienkowski MJ;  
 PI Heintz RL;  
 XX WPI; 2002-216995/27.  
 XX Novel substrates for human aspartyl protease useful for identifying  
 PT modulators of beta secretase activity of aspartyl protease for treating  
 PT Alzheimer's disease.  
 XX Disclosure; Page 188; 188pp; English.  
 XX The present invention describes an isolated peptide (I) comprising a  
 CC sequence of at least four amino acids, where the peptide is a substrate  
 CC for conducting aspartyl protease assays. (I) has neuroprotective and  
 CC nontropic activities, and can be used as an inhibitor of beta-secretase  
 CC activity. A beta-secretase modulator from the present invention can be  
 CC used for inhibiting beta-secretase activity in vivo, and in the  
 CC manufacture of a medicament for the treatment of Alzheimer's disease.  
 CC Pharmaceutical compositions from the present invention can be used for  
 CC treating a disease or condition characterised by an abnormal beta-  
 CC secretase activity. (I) is useful for identifying agents that modulate  
 CC the activity of human Asp2 aspartyl protease (Hu-Asp2). (I) is useful as  
 CC a core structure to construct derivatives. ABL49914 to ABL49925 and  
 CC ABB06409 to ABB06593 represent sequences used in the exemplification of  
 CC the present invention  
 XX Sequence 12 AA;  
 SQ Query Match 95.0%; Score 19; DB 5; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 DB 5 VAEF 8

DT 15-NOV-2002 (first entry)  
 XX Memapsin 2 substrate specificity determination peptide #10.  
 DS Human; memapsin 2; aspartic protease; beta secretase;  
 XX degenerative disease; Alzheimer's disease; amyloid precursor protein;  
 KW APP; neuroprotective; nontropic; inhibitor;  
 KW substrate side-chain preference.  
 XX Homo sapiens.  
 OS Synthetic.  
 XX WO200253594-A2.  
 PN 11-JUL-2002.  
 PD 28-DEC-2001; 2001WO-US050826.  
 PF 28-DEC-2000; 2000US-0258705P.  
 PX 14-MAR-2001; 2001US-0275756P.  
 PR (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 XX (UNII) UNIV ILLINOIS FOUND.  
 PA Tang JN, Koelsch G, Ghosh AK;  
 PX WPI; 2002-619088/66.  
 PT New memapsin 2 activity inhibitor useful in treatment of e.g. Alzheimer's  
 PT disease.  
 XX Example 5; Page 51; 74pp; English.  
 XX The invention relates to an inhibitor of catalytically active memapsin 2  
 CC (an aspartic protease which can cleave at beta secretase sites), which  
 CC binds to the active site of memapsin 2 defined by the presence of two  
 CC catalytic aspartic residues and substrate binding cleft. Also included is  
 CC a method of determination of the substrate side-chain preference in  
 CC memapsin 2 sub-sites comprising: (a) reacting a mixture of memapsin 2  
 CC substrates with memapsin 2, and determining the sub-site preference of  
 CC memapsin 2 by determining relative initial hydrolysis rates of the  
 CC mixture of memapsin 2 substrates; or (b) preparing a combinatorial  
 CC library of memapsin 2 inhibitors containing a base sequence taken from  
 CC OM99-2 (Glu-Val-Asn-Leu-Ala-Glu-phe), probing the library of  
 CC inhibitors with memapsin 2 which binds to several inhibitors to generate  
 CC several bound memapsin 2, and detecting the bound memapsin 2 with an  
 CC antibody raised to memapsin 2 and an alkaline phosphatase conjugated  
 CC secondary antibody. The inhibitors may be used in the manufacture of a  
 CC medicament for the treatment of Alzheimer's disease since memapsin 2 may  
 CC be involved in the cleavage of amyloid precursor protein (APP), and for  
 CC determining the substrate side-chain preference in memapsin 2 sub-sites.  
 CC The present sequence represents a sub-site variant peptide used to  
 CC determine the substrate specificity of human memapsin 2  
 XX Sequence 12 AA;  
 SQ Query Match 95.0%; Score 19; DB 5; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 DB 5 VAEF 8  
 RESULT 14  
 ADD35467  
 ID ADD35467 standard; peptide; 12 AA.  
 XX ADD35467;  
 XX 15-JAN-2004 (first entry)  
 DT 15-JAN-2004 (first entry)  
 XX

DE Escherichia coli DnaK peptide #3.  
 XX crystallised recombinant protein; metabolism; Staphylococcus aureus;  
 KW Streptococcus pneumoniae; Helicobacter pylori; Escherichia coli;  
 KW Pseudomonas aeruginosa; vaccine.  
 XX  
 OS Escherichia coli.  
 XX  
 PN WO2003044185-A2.  
 XX  
 XX 30-MAY-2003.  
 XX  
 XX 21-NOV-2002; 2002WO-CA001768.  
 XX  
 XX 21-NOV-2001; 2001US-0332160P.  
 PR 21-NOV-2001; 2001US-0333661P.  
 PR 27-NOV-2001; 2001US-0333665P.  
 PR 18-DEC-2001; 2001US-0341770P.  
 PR 19-DEC-2001; 2001US-0341954P.  
 PR 19-DEC-2001; 2001US-0342003P.  
 PR 20-DEC-2001; 2001US-0342542P.  
 PR 21-DEC-2001; 2001US-0344252P.  
 PR 28-DEC-2001; 2001US-0343570P.  
 PR 28-DEC-2001; 2001US-0343606P.  
 PR 28-DEC-2001; 2001US-0343679P.  
 XX  
 PA (AFFI-) AFFINIUM PHARM INC.  
 XX  
 XX Edwards A, Dharami A, Vedadi M, Alan MZ, Awrey D, Beattie B;  
 PI Canadien V, Domagala M, Houston S, Mansoury K, Necakov S, Nethery K;  
 PI Ng I, Pinder B, Sheidrick B, Vallee F, Wrezel O;  
 XX WPI; 2003-513596/48.  
 XX  
 XX New crystallized recombinant polypeptides from Staphylococcus aureus,  
 PT Streptococcus pneumoniae, Helicobacter pylori or Pseudomonas aeruginosa  
 PT involved in general metabolism, useful as drug targets for pathogenic  
 PT bacteria.  
 XX  
 XX Disclosure; SEQ ID NO 66; 277pp; English.  
 PS  
 XX The invention comprises a crystallised recombinant protein that is  
 CC involved in general metabolism, the recombinant protein may be from  
 CC Staphylococcus aureus, Streptococcus pneumoniae, Helicobacter pylori,  
 CC Escherichia coli or Pseudomonas aeruginosa. The crystallised recombinant  
 CC protein of the invention is useful in the prevention (vaccine) or  
 CC treatment of a disease or disorder caused by S. pneumoniae, H. pylori, E.  
 CC coli or P. aeruginosa. The present amino acid sequence was used in the  
 CC exemplification of the invention.  
 XX  
 SQ Sequence 12 AA;  
 XX  
 XX Query Match 95.0%; Score 19; DB 7; Length 12;  
 XX Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 2 VAEF 5  
 DB 9 VAEF 12  
 XX  
 XX RESULT 15  
 XX AAB07889  
 ID AAB07889 standard; peptide; 13 AA.  
 XX  
 XX AAB07889;  
 AC  
 XX  
 XX 14-NOV-2000 (first entry)  
 DT  
 XX  
 DE A beta-secretase inhibitor peptide.  
 XX  
 XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;

KW inhibitor.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200047618-A2.  
 XX  
 XX 17-AUG-2000.  
 XX  
 XX 10-FEB-2000; 2000WO-US003819.  
 PF  
 XX 10-FEB-1999; 99US-0119571P.  
 PR 15-JUN-1999; 99US-0139172P.  
 PR  
 XX (ELAN-) ELAN PHARM INC.  
 PA  
 XX Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 PI Sinha S, Tateuno G, Tung J, Wang S, Mcconlogue L;  
 XX WPI; 2000-533011/48.  
 DR  
 XX Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease.  
 PT  
 XX Claim 33; Page 24; 121pp; English.  
 PS  
 XX The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to a  
 CC mammalian subject e.g. with Alzheimer's disease or Alzheimer's disease-  
 CC like pathology to test if they maintain or improve cognitive ability or  
 CC reduce the plaque burden. The compounds are used for the treatment of  
 CC amyloidogenic diseases e.g. Alzheimer's disease. The present sequence  
 CC represents an inhibitor of beta-secretase enzyme  
 XX  
 SQ Sequence 13 AA;  
 XX  
 XX Query Match 95.0%; Score 19; DB 3; Length 13;  
 XX Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 2 VAEF 5  
 DB 10 VAEF 13  
 XX  
 XX RESULT 16  
 XX AAM99276  
 ID AAM99276 standard; peptide; 13 AA.  
 XX  
 XX AAM99276;  
 AC  
 XX 07-DEC-2001 (first entry)  
 DT  
 XX  
 DE Vaccine related MHC ligand peptide SEQ ID NO:379.  
 XX  
 XX Glutamic acid; glutamine; vaccine; major histocompatibility complex; MHC;  
 KW immunomodulator; antiallergic; endocrine; neuroprotectant; virucidal;  
 KW bactericidal; antiparasitic; fungicidal; cytostatic; medicine;  
 KW pharmaceutical; immune disorder; immune deficiency; autoimmune;  
 KW hypersensitivity; allergy; graft rejection; infection; hormonal disorder;  
 KW central nervous system disease; cancer; melanoma; anti-melanoma vaccine;  
 KW human immunodeficiency virus.  
 XX  
 XX Mycobacterium leprae.  
 OS  
 XX WO200170772-A2.  
 XX  
 XX 27-SBP-2001.  
 PD  
 XX 22-MAR-2001; 2001WO-FR000872.  
 PF

XX 23-MAR-2000; 2000FR-00003711.  
 XX (FABR ) FABRE MEDICAMENT SA PIERRE.  
 XX KLinguer-Hamour C, Corvaia N, Beck A, Goetsch L;  
 XX WPI; 2001-611470/70.  
 XX Stabilized pharmaceutical containing N-terminal glutamic acid or  
 PT glutamine, useful e.g. in anti-melanoma vaccines, is an addition salt  
 PT with strong acid.  
 XX Claim 9; Page 96; 149pp; French.  
 XX The present invention describes a pharmaceutical compound (I) that  
 CC contains an N-terminal glutamic acid (Glu) or glutamine (Gln) residue in  
 CC the form of an addition salt with a strong, physiologically acceptable  
 CC acid (II). Also described are: (a) a pharmaceutical composition  
 CC containing at least one (I); (b) a vaccine containing at least one (I)  
 CC where this is a major histocompatibility complex (MHC) ligand (Ia); (c) a  
 CC method for in vitro diagnosis of diseases associated with the presence of  
 CC (Ia); (d) a kit for method (c) that includes a (Ia); and (e) a process  
 CC for preparing (I). (I) has immunomodulator, endocrine, antiallergic,  
 CC neuroprotectant, virucidal, bactericidal, antiparasitic, fungicidal and  
 CC cytostatic activities. (I) are useful, in human or veterinary medicine,  
 CC in pharmaceutical compositions (for treating immune disorders, e.g.  
 CC immune deficiency, autoimmune states, hypersensitivity, allergy, graft  
 CC rejection, infection, hormonal disorders and central nervous system  
 CC diseases), also, where (I) is a MHC ligand (Ia), in vaccines for  
 CC treatment or prevention of: (i) viral, bacterial, parasitic or fungal  
 CC infections; or (ii) of cancers. A particular application is in anti-  
 CC melanoma vaccines. (I) are also useful for in vitro diagnosis of diseases  
 CC associated with interactions between MHC and (I), e.g. melanoma and human  
 CC immunodeficiency virus infection. AAM93898 to AAM9392 represent peptides  
 CC which can be used in pharmaceutical compounds from the present invention  
 XX  
 SQ Sequence 13 AA;  
 Query Match 95.0%; Score 19; DB 4; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 DB 4 VAEF 7  
 RESULT 17  
 AAB07888  
 ID AAB07888 standard; peptide; 14 AA.  
 XX AAB07888;  
 AC  
 XX 14-NOV-2000 (first entry)  
 DT  
 XX A peptide fragment derived from beta-amyloid precursor protein.  
 DE  
 XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 KW inhibitor.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO2000047618-A2.  
 FN  
 PD 17-AUG-2000.  
 PD  
 XX 10-FEB-2000; 2000WO-US003819.  
 PF  
 XX 10-FEB-1999; 99US-0119571P.  
 PR 15-JUN-1999; 99US-0139172P.  
 XX

PA (ELAN-) ELAN PHARM INC.  
 XX Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
 XX WPI; 2000-533011/48.  
 XX Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease.  
 XX Disclosure; Page 12; 121pp; English.  
 XX The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to a  
 CC mammalian subject e.g. with Alzheimer's disease or Alzheimer's disease-  
 CC like pathology to test if they maintain or improve cognitive ability or  
 CC reduce the plaque burden. The compounds are used for the treatment of  
 CC amyloidogenic diseases e.g. Alzheimer's disease. The present sequence  
 CC represents a peptide derived from beta-amyloid precursor protein  
 XX Sequence 14 AA;  
 SQ  
 Query Match 95.0%; Score 19; DB 3; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 DB 11 VAEF 14  
 RESULT 18  
 AAM97460  
 ID AAM97460 standard; peptide; 14 AA.  
 XX  
 AC AAM97460;  
 XX 24-JAN-2002 (first entry)  
 DT  
 XX Human peptide #735 encoded by a SNP oligonucleotide.  
 DE  
 XX Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;  
 KW neuroprotective; antimicrobial; gene therapy; vaccine; amyase; cancer;  
 KW amyloid protein; angiopoietin; apoptosis related protein; cadherin;  
 KW cyclin; polymerase; oncogene; kinase; colony stimulating factor;  
 KW complement related protein; cytochrome; kinesin; cytokine; interferon;  
 KW interleukin; G-protein coupled receptor; thioesterase; inflammation;  
 KW multifactorial disease; autoimmune disease; infection;  
 KW nervous system disease.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200147944-A2.  
 FN  
 XX 05-JUL-2001.  
 PD  
 XX 28-DEC-2000; 2000WO-US035498.  
 PF  
 XX 28-DEC-1999; 99US-0173419P.  
 PR 27-DEC-2000; 2000US-00173419.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX Shimkets RA, Leach M;  
 PI  
 XX WPI; 2001-465210/50.  
 DR  
 XX Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,  
 PT oncogenes and histones, useful for diagnosing and treating, e.g. cancer,  
 PT

PT autoimmune diseases and infections.

PS Disclosure: Page 3829: 4143pp: English.

XX The present invention relates to oligonucleotides (see AAL26793-AAL34659)  
CC encoding polymorphic variants of proteins related to amyloses, amyloid  
CC proteins, angiotensin, apoptosis related proteins, cadherin, cyclin,  
CC polymerase, oncogenes, histones, kinases, colony stimulating factors,  
CC complement related proteins, cytochromes, kinases, cytokines,  
CC interferons, interleukins, G-protein coupled receptors and thioesterases.  
CC The present sequence is a peptide encoded by one such oligonucleotide.  
CC The oligonucleotides and the peptides encoded by them may be used in the  
CC prevention, diagnosis and treatment of diseases associated with  
CC inappropriate expression of the proteins listed above. Disorders that may  
CC be prevented, diagnosed and/or treated include multifactorial diseases  
CC with a genetic component, such as autoimmune diseases (e.g. rheumatoid  
CC arthritis, multiple sclerosis, diabetes, systemic lupus erythematosus  
CC and Grave's disease), inflammation, cancer (e.g. cancers of the bladder,  
CC brain, breast, colon and kidney, leukaemia), diseases of the nervous  
CC system and an infection of pathogenic organisms  
XX  
XX Sequence 14 AA:  
SO

Query Match	95.0%;	Score 19;	DB 4;	Length 14;
Best Local Similarity	100.0%;	Pred. No. 1.8e+02;		
Matches 4;	Conservative	0;	Mismatches 0;	Indels 0;
Gaps 0;				

Qy	2	VAEF	5
Db	5	VAEF	8

RESULT 19  
AAE32223  
ID AAE32223 standard; peptide; 15 AA.

AA AAE32223;  
AC

XX  
DT 24-MAR-2003 (first entry)

XX DE Arabidopsis thaliana S11 peptide #6.

Genomic database; mass spectrometer; proteomic business; pharmaceutical;  
 nuclear transport; signalling pathway; cellular organelle.

XX OS Arabidopsis thaliana.

XX  
PN  
WO200280649-A2.17-OCT-2002-  
PD  
XXXX  
PF 09-APR-2002; 2002WO-US011417.XX  
PR 09-APR-2001: 2001US-0282551P.

PR 20-APR-2001; 2001US-028536ZF.  
XX

PA (MDSP-) MDS PROTEOMICS INC.  
XX

PI Mann M, Mortensen P;  
XX

DR WPI; 2003-058452/05.  
XX

PT Identifying coding sequence in genomic databases for conducting  
PT proteomics business, by using polypeptide sequence information obtained  
PT from peptide sequencing projects, especially those using mass  
PT spectrometers.

xx PS Example 1: Page 41; 100pp; English.

xx The invention relates to a method for identifying coding sequence in  
 CC genomic databases, by using polypeptide sequence information obtained  
 CC from peptide sequencing projects, especially those using mass  
 CC spectrometers. It is useful for conducting a proteomics business. It is

CC also used for establishing a distribution system for distributing the  
CC pharmaceutical preparation for sale, and may optionally include  
CC establishing a sales group for marketing the pharmaceutical preparation.  
CC It is also used for predicting the gene structure, such as intron/exon  
CC boundaries, for searching genomic databases for sequences derived from  
CC multi-protein complexes e.g. assemblies with a particular function such  
CC as splicing, transport or nuclear import or export; for elucidating  
CC transient rather than structural complexes (that are involved in  
CC signalling pathways) and for identifying proteins in cellular organelles.  
CC The present sequence is Arabidopsis thaliana peptide used to illustrate  
CC the method of the invention  
CC  
CC  
CC Sequence 15 AA;  
SQ

Query Match	95.0%;	Score 19;	DB 6;	Length 15;
Best Local Similarity	100.0%;	Pred. No. 1.9e+02;		
Matches	4;	Conservative 0;	Mismatches 0;	Indels 0;
		Gaps 0;		

Qy	2	5
	VAEF	VAEF
Db	10	13
	VAEF	VAEF

RESULT 20  
AAW42944  
ID AAW42944 standard; peptide; 20 AA.

AAW42944;

DT 28-APR-1998 (first entry)

Immunogenic Hepatitis A virus peptide YK-1369.

Immunogenic peptide; immunogenic epitope; P3A protein; immune response; antibody.  
KW

OS Synthetic.

XX  
OS  
неприятель А. В. Л.

XX  
XX  
NA 4105/60M 4105/60M

FD-302 (Rev. 1-25-60)

XX  
BT  
T8-APR-1997; 5/MO-US0006891.

FR 19-APR-1996; 96US-0015644P.  
XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES:  
XX

PI  
XX  
Fields HA, Khudyakov YE;

DR WPI; 1997-535831/49.  
XX

PT response to HAV in a mammal or to detect the presence of antibodies against HAV in a mammal.

XX  
ps  
Claim 33: Page 115: 140pp: English:

Peptides AAW42943-46 are immunogenic peptides corresponding to immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are substantially similar to a portion of the amino acid sequence of the p3A protein of HAV corresponding to amino acids 1423-1496. The present peptide is derived from amino acids 1430-1449 and has a reactivity of 58.5% with acute sera. Compositions containing the peptides can be used to induce an immune response to HAV in a mammal. The peptides can also be used to detect the presence of antibodies against HAV in mammalian serum. The peptides can also be used to make an antibody against HAV by administering the peptide to a mammal.

Sequence: 20 AA;

Query Match: 95.0%; Score 19; DB 2; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.6e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0;

Y 2 VAEF 5  
b 4 VAEF 7

RESULT 21  
AAW42943  
D AAW42943 standard; peptide; 20 AA.  
X  
C AAW42943;  
X  
X 28-APR-1998 (first entry)  
X  
X Immunogenic Hepatitis A virus peptide YK-1368.  
X  
X Immunogenic peptide; immunogenic epitope; P3A protein; immune response;  
X antibody.  
X Synthetic.  
X Hepatitis A virus.  
X WO9740147-A1.  
X  
X 30-OCT-1997.  
X  
X 18-APR-1997; 97WO-US006891.  
X  
X 19-APR-1996; 96US-0015644P.  
X (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
X Fields HA, Khudyakov YE;  
X WPI; 1997-535831/49.  
X Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
X response to HAV in a mammal or to detect the presence of antibodies  
X against HAV in a mammal.  
X Claim 33; Page 115; 140pp; English.  
X Peptides AAW42943-46 are immunogenic peptides corresponding to  
X immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
X substantially similar to a portion of the amino acid sequence of the P3A  
X protein of HAV corresponding to amino acids 1423-1496. The present  
X peptide is derived from amino acids 1421-1440 and has a reactivity of  
X 63.4% with acute sera. Compositions containing the peptides can be used  
X to induce an immune response to HAV in a mammal. The peptides can also be  
X used to detect the presence of antibodies against HAV in mammalian serum.  
X The peptides can also be used to make an antibody against HAV by  
X administering the peptide to a mammal

Sequence 20 AA;  
Query Match 95.0%; Score 19; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5  
b 13 VAEF 16

RESULT 22  
AAB69462  
D AAB69462 standard; peptide; 21 AA.  
X  
X AAB69462;  
X  
X 20-APR-2001 (first entry)

XX Synthetic HAV P3A peptide, SEQ ID NO: 62.  
DE Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
XW antigen; major structural capsid polypeptide; HAV antibody detection.  
XX Hepatitis A virus.  
OS Synthetic.  
XW WO200105824-A2.  
EN 25-JAN-2001.  
XX 14-JUL-2000; 2000MO-US019267.  
PF 15-JUL-1999; 99US-0144412P.  
XW (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX Fields HA, Khudyakov YE;  
XW WPI; 2001-112681/12.  
DR Synthetic peptides used as antigen sources for enzyme immunoassays  
XW detecting anti-hepatitis A virus and as vaccines.  
XW Claim 22; Page 108; 130pp; English.  
XW The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IGM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy end  
CC of the peptides enhances the IGM antibody reactivity

Sequence 21 AA;  
Query Match 95.0%; Score 19; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0;

Qy 2 VAEF 5  
b 13 VAEF 16

RESULT 23  
AAB69463  
D AAB69463 standard; peptide; 21 AA.  
X  
X AAB69463;  
X  
X 20-APR-2001 (first entry)  
X  
X Synthetic HAV P3A peptide, SEQ ID NO: 63.  
DE Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
XW antigen; major structural capsid polypeptide; HAV antibody detection.  
XX Hepatitis A virus.  
OS Synthetic.  
XW WO200105824-A2.

XX PD 25-JAN-2001.  
 XX PF 14-JUL-2000; 2000WO-US019267.  
 XX PR 15-JUL-1999; 99US-0144412P.  
 XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX PI Fields HA, Khudyakov YE;  
 XX PT WPI; 2001-112681/12.  
 XX PR Synthetic peptides used as antigen sources for enzyme immunoassays  
 XX PT detecting anti-hepatitis A virus and as vaccines.  
 XX PS Claim 22; Page 109; 130pp; English.  
 XX CC The present sequence is one of a number of synthetic peptides which are  
 CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
 CC comprise antigenic epitopes of the major structural capsid polypeptides  
 CC or non-structural polypeptides of HAV with one or more glutamine  
 CC molecules at the carboxy end of the peptide. The peptides are used to  
 CC detect the presence of antibodies against HAV in mammalian serum, to  
 CC detect the presence of HAV in a human or animal through the binding of  
 CC the peptide to an antibody, to detect acute phase infection by detecting  
 CC IGM antibodies in mammalian serum and detecting convalescence in a  
 CC mammal. The peptides are used to detect or quantify HAV antibodies in  
 CC samples in clinical or research-based assays using immunoblotting,  
 CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
 CC tracking of radioactive or bioluminescent markers, chromatography or  
 CC electrophoresis. The peptides are used to induce an immune response to  
 CC HAV when administered to a human or animal. Glutamine at the carboxy end  
 CC of the peptides enhances the IGM antibody reactivity  
 XX SQ Sequence 21 AA;  
 Query Match 95.0%; Score 19; DB 4; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 Db ||||  
 4 VAEF 7  
 RESULT 24  
 ID AAW42946  
 AC AAW42946; peptide; 25 AA.  
 AC AAW42946;  
 XX 28-APR-1998 (first entry)  
 XX Immunogenic Hepatitis A virus peptide YK-1832.  
 XX Immunogenic peptide; immunogenic epitope; P3A protein; immune response;  
 XX antibody.  
 XX Synthetic.  
 XX Hepatitis A virus.  
 XX WO9740147-A1.  
 XX 30-OCT-1997.  
 XX 18-APR-1997; 97WO-US006891.  
 XX 19-APR-1996; 96US-0015644P.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Fields HA, Khudyakov YE;

XX WPI; 1997-535831/49.  
 XX Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
 XX response to HAV in a mammal or to detect the presence of antibodies  
 XX against HAV in a mammal.  
 XX PS Claim 33; Page 115; 140pp; English.  
 XX CC The present immunogenic peptide corresponds to an immunogenic epitope of  
 CC the Hepatitis A virus (HAV). The peptide is substantially similar to a  
 CC portion of the amino acid sequence of the P3A protein of HAV  
 CC corresponding to amino acids 1423-1496. Compositions containing the  
 CC peptide can be used to induce an immune response to HAV in a mammal. The  
 CC peptide can also be used to detect the presence of antibodies against HAV  
 CC in mammalian serum. The peptide can also be used to make an antibody  
 CC against HAV by administering the peptide to a mammal  
 XX SQ Sequence 25 AA;  
 Query Match 95.0%; Score 19; DB 2; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 Db ||||  
 3 VAEF 6  
 RESULT 25  
 ID AAB69465  
 AC AAB69465 standard; peptide; 25 AA.  
 XX AAB69465;  
 XX 20-APR-2001 (first entry)  
 XX Synthetic HAV P3A peptide, SEQ ID NO: 65.  
 XX Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
 XX antigen; major structural capsid polypeptide; HAV antibody detection.  
 XX Hepatitis A virus.  
 XX Synthetic.  
 XX WO200105824-A2.  
 XX 25-JAN-2001.  
 XX 14-JUL-2000; 2000WO-US019267.  
 XX 15-JUL-1999; 99US-0144412P.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Fields HA, Khudyakov YE;  
 XX WPI; 2001-112681/12.  
 XX Synthetic peptides used as antigen sources for enzyme immunoassays  
 XX detecting anti-hepatitis A virus and as vaccines.  
 XX Claim 22; Page 110; 130pp; English.  
 XX The present sequence is one of a number of synthetic peptides which are  
 XX immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
 XX comprise antigenic epitopes of the major structural capsid polypeptides  
 XX or non-structural polypeptides of HAV with one or more glutamine  
 XX molecules at the carboxy end of the peptide. The peptides are used to  
 XX detect the presence of antibodies against HAV in mammalian serum, to  
 XX detect the presence of HAV in a human or animal through the binding of  
 XX the peptide to an antibody, to detect acute phase infection by detecting  
 XX IGM antibodies in mammalian serum and detecting convalescence in a  
 XX mammal. The peptides are used to detect or quantify HAV antibodies in  
 XX samples in clinical or research-based assays using immunoblotting,  
 XX fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
 XX tracking of radioactive or bioluminescent markers, chromatography or  
 XX electrophoresis. The peptides are used to induce an immune response to  
 XX HAV when administered to a human or animal. Glutamine at the carboxy end  
 XX of the peptides enhances the IGM antibody reactivity

C mammal. The peptides are used to detect or quantify HAV antibodies in  
C samples in clinical or research-based assays using immunoblotting,  
C fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
C tracking of radioactive or bioluminescent markers, chromatography or  
C electrophoresis. The peptides are used to induce an immune response to  
C HAV when administered to a human or animal. Glutamine at the carboxy end  
C of the peptides enhances the IGM antibody reactivity  
X  
Q Sequence 25 AA;

Query Match 95.0%; Score 19; DB 4; Length 25;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5  
b 3 VAEF 6

RESULT 26  
AAW17832  
ID AAW17832 standard; peptide; 28 AA.

AC AAW17832;

AT 29-JUL-1997 (first entry)

BE Human mevalonate pyrophosphate decarboxylase peptide NT71.

GW Mevalonate pyrophosphate decarboxylase; MPD; cholesterol.

JS Homo sapiens.

PN WO9714787-A1.

PD 24-APR-1997.

PF 10-OCT-1996; 96WO-EP004394.

PR 18-OCT-1995; 95US-0005652P.

PA (NOVS ) NOVARTIS AG.

PI Toth MJ, Huwylar LR;

DR WPI; 1997-245104/22.

PT Human mevalonate pyrophosphate decarboxylase coding sequence - used for  
PT screening for MPD inhibitors, which regulate and control cholesterol  
PT synthesis.

PS Example 3; Page 10; 37pp; English.

CC Peptide NT71 (AAW17832) was obtd. by trypsin digestion of human liver  
CC mevalonate pyrophosphate decarboxylase (MPD) (AAW17831), an enzyme of the  
CC cholesterol biosynthetic pathway. The sequence of the peptide was used to  
CC design PCR primers utilised in the amplification of cDNA from a rat liver  
CC cDNA library. A rat MPD partial clone was obtd. and used as a probe to  
CC screen a human liver cDNA library. A 1800 bp sequence (AAW66464) coding  
CC for human MPD was identified

XX Sequence 28 AA;

Query Match 95.0%; Score 19; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
DB 20 VAEF 23

RESULT 27

AAW33910  
ID AAW33910 standard; protein; 28 AA.  
XX AC AAW33910;  
XX AC AAW33910;  
XX DT 17-OCT-2001 (first entry)

XX Peptide #7947 encoded by probe for measuring placental gene expression.  
DE XX  
DE XX  
DE XX  
KW Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.

OS Homo sapiens.

XX WO200157272-A2.

FN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000663.

XX 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human placenta.

XX Claim 27; SEQ ID NO 34179; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENP:  
XX see AAI3315-AAI57546). The present sequence is a peptide encoded by one  
XX such probe. The probes are useful for producing a microarray for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from human placenta. The probes are useful for antenatal diagnosis of  
XX human genetic disorders

XX Sequence 28 AA;

Query Match 95.0%; Score 19; DB 4; Length 28;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
DB 4 VAEF 7

RESULT 28  
ABG55464  
ID ABG55464 standard; peptide; 28 AA.

XX AC ABG55464;

XX DT 25-FEB-2003 (first entry)

XX Human liver peptide, SEQ ID No 34112.

XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
XX hypercholesterolaemia; coronary heart disease.

XX Homo sapiens.

XX WO200157273-A2.

XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US000664.  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00623366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-48898/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human adult liver.  
XX Claim 27; SEQ ID NO 34112; 658pp; English.  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
XX measuring human gene expression in a sample derived from human adult  
XX liver, comprising one of 13109 defined nucleotide sequences given in the  
XX specification (or complements/fragments). The probe hybridises at high  
XX stringency to a nucleic acid molecule expressed in the human adult liver.  
XX (I) may be used for predicting, measuring and displaying gene expression  
XX in samples derived from human adult liver. The genes identified may be  
XX involved in genetic liver diseases such as cirrhosis,  
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
XX associated with coronary heart disease. ABG47348-ABG5930 represent human  
XX liver single exon encoded peptides of the invention. Note: The sequence  
XX information for this patent does not appear in the printed specification  
XX but was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 28 AA;  
XX  
XX Query Match 95.0%; Score 19; DB 4; Length 28;  
XX Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 2 VAEF 5  
XX |||||  
XX 4 VAEF 7  
XX  
XX DB  
XX  
XX RESULT 29  
XX ABG43601  
XX ID ABG43601 standard; peptide; 28 AA.  
XX  
XX AC ABG43601;  
XX  
XX DT 19-AUG-2002 (first entry)  
XX  
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 33266.  
XX  
XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
XX chronic obstructive pulmonary disease; interstitial lung disease;  
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;  
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;  
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
XX primary ciliary dyskinesia; pulmonary hypertension;  
XX hyaline membrane disease.  
XX  
XX OS Homo sapiens.  
XX  
XX FN WO200186003-A2.

XX 15-NOV-2001.  
XX 30-JAN-2001; 2001WO-US000665.  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00623366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
XX measure gene expression in human lung samples.  
XX Claim 27; SEQ ID NO 33266; 634pp; English.  
XX The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human lung comprising single exon nucleic acid probes having one of  
XX 12614 nucleic acid sequences mentioned in the specification, or their  
XX complements or the 12387 open reading frames derived from the 12614  
XX probes. Also included are a microarray comprising the novel set of probes  
XX; the novel set of probes which hybridise at high stringency to a nucleic  
XX acid expressed in the human lung; measuring gene expression in a sample  
XX derived from human lung, comprising (a) contacting the array with a  
XX collection of detectably labeled nucleic acids derived from human lung  
XX mRNA, and (b) measuring the label detectably bound to each probe of the  
XX array; identifying exons in a eukaryotic genome, comprising (a)  
XX algorithmically predicting at least one exon from genomic sequences of  
XX the eukaryote; and (b) detecting specific hybridisation of detectably  
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
XX having a fragment identical to the predicted exon, the probe is included  
XX in the above mentioned microarray; assigning exons to a single gene,  
XX comprising (a) identifying exons from genomic sequence by the method  
XX above and (b) measuring the expression of each of the exons in several  
XX tissues and/or cell types using hybridisation to a single exon  
XX microarrays having a probe with the exon, where a common pattern of  
XX expression of the exons in the tissues and/or cell types indicates that  
XX the exons should be assigned to a single gene; a peptide comprising one  
XX of 12011 sequences, mentioned in the specification, or encoded by the  
XX probes/open reading frames (ORF). The probes are used for gene expression  
XX analysis, and for identifying exons in a gene, particularly using human  
XX lung derived mRNA and for the study of lung diseases such as asthma, lung  
XX cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
XX disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
XX tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
XX Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
XX histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
XX Karsenger syndrome, fibrocystic pulmonary fibrosis, primary ciliary  
XX dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
XX present sequence is a peptide/protein encoded by a single exon probe of  
XX the invention. Note: The sequence data for this patent did not form part  
XX of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 28 AA;  
XX  
XX Query Match 95.0%; Score 19; DB 5; Length 28;  
XX Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 2 VAEF 5  
XX |||||  
XX 4 VAEF 7  
XX  
XX DB

```

RESULT 30
AAW62760
D AAW62760 standard; protein; 30 AA.
X
C AAW62760;
X
T 09-NOV-1998 (first entry)
X Streptococcus pneumoniae polypeptide.
X Polypeptide; ORF; open reading frame; infection; bacterial;
X streptococcal; bacteremia; diagnosis; prophylaxis.
X Streptococcus pneumoniae.
X WO9823631-A1.
X
D 04-JUN-1998.
X
F 24-NOV-1997; 97WO-US021976.
X
R 27-NOV-1996; 96US-0031879P.
X (SMIK ) SMITHKLINE BEECHAM CORP.
X (SMIK ) SMITHKLINE BEECHAM PLC.
X Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
X Reid RH, Zarfos PN;
X WPI; 1998-322654/28.
X Streptococcus pneumoniae polynucleotides - useful for developing products
X for diagnosis, prevention and treatment of infections e.g. pneumonia,
X bacteremia, meningitis or endocarditis.
X Claim 5; Page 32; 181pp; English.
X The sequence is that of a Streptococcal polypeptide. The polypeptide can
X potentially be used for the diagnosis and prevention of bacterial
X infections, especially SP infection. It may be used for the treatment of
X diseases such as otitis media, conjunctivitis, pneumonia, bacteremia,
X meningitis, sinusitis, pleural empyema, endocarditis or infection of the
X cerebrospinal fluid
X
X Sequence 30 AA;
Query Match 95.0%; Score 19; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
Db 21 VAEF 24

RESULT 31
AAW82654
ID AAW82654 standard; protein; 32 AA.
X
AC AAW82654;
X
DT 07-NOV-2001 (first entry)
X Human immune/haematopoietic antigen SEQ ID NO:10247.
X Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
X cytostatic; gene therapy; vaccine; metastasis.
X Homo sapiens.
X WO200157182-A2.
X

```

```

PD 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WC-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184664P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189874P.
XX
XX 17-MAR-2000; 2000US-0190076P.
XX
XX 18-APR-2000; 2000US-0198123P.
XX
XX 19-MAY-2000; 2000US-0205515P.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX
XX 28-JUN-2000; 2000US-0214886P.
XX
XX 30-JUN-2000; 2000US-0215135P.
XX
XX 07-JUL-2000; 2000US-0216647P.
XX
XX 07-JUL-2000; 2000US-0216880P.
XX
XX 11-JUL-2000; 2000US-0217487P.
XX
XX 14-JUL-2000; 2000US-0217496P.
XX
XX 26-JUL-2000; 2000US-0220963P.
XX
XX 26-JUL-2000; 2000US-0220964P.
XX
XX 14-AUG-2000; 2000US-0224518P.
XX
XX 14-AUG-2000; 2000US-0224519P.
XX
XX 14-AUG-2000; 2000US-0225213P.
XX
XX 14-AUG-2000; 2000US-0225214P.
XX
XX 14-AUG-2000; 2000US-0225266P.
XX
XX 14-AUG-2000; 2000US-0225267P.
XX
XX 14-AUG-2000; 2000US-0225268P.
XX
XX 14-AUG-2000; 2000US-0225270P.
XX
XX 14-AUG-2000; 2000US-0225447P.
XX
XX 14-AUG-2000; 2000US-0225757P.
XX
XX 14-AUG-2000; 2000US-0225758P.
XX
XX 14-AUG-2000; 2000US-0225759P.
XX
XX 18-AUG-2000; 2000US-0226279P.
XX
XX 22-AUG-2000; 2000US-0226681P.
XX
XX 22-AUG-2000; 2000US-0226868P.
XX
XX 22-AUG-2000; 2000US-0227182P.
XX
XX 23-AUG-2000; 2000US-0227009P.
XX
XX 30-AUG-2000; 2000US-0228924P.
XX
XX 01-SEP-2000; 2000US-0229287P.
XX
XX 01-SEP-2000; 2000US-0229343P.
XX
XX 01-SEP-2000; 2000US-0229344P.
XX
XX 01-SEP-2000; 2000US-0229345P.
XX
XX 05-SEP-2000; 2000US-0229509P.
XX
XX 05-SEP-2000; 2000US-0229513P.
XX
XX 06-SEP-2000; 2000US-0230437P.
XX
XX 08-SEP-2000; 2000US-0230438P.
XX
XX 08-SEP-2000; 2000US-0231242P.
XX
XX 08-SEP-2000; 2000US-0231243P.
XX
XX 08-SEP-2000; 2000US-0231244P.
XX
XX 08-SEP-2000; 2000US-0231413P.
XX
XX 08-SEP-2000; 2000US-0231414P.
XX
XX 08-SEP-2000; 2000US-0232080P.
XX
XX 08-SEP-2000; 2000US-0232081P.
XX
XX 12-SEP-2000; 2000US-0231968P.
XX
XX 14-SEP-2000; 2000US-0232397P.
XX
XX 14-SEP-2000; 2000US-0232398P.
XX
XX 14-SEP-2000; 2000US-0232399P.
XX
XX 14-SEP-2000; 2000US-0232400P.
XX
XX 14-SEP-2000; 2000US-0232401P.
XX
XX 14-SEP-2000; 2000US-0233063P.
XX
XX 14-SEP-2000; 2000US-0233064P.
XX
XX 14-SEP-2000; 2000US-0233065P.
XX
XX 21-SEP-2000; 2000US-0234223P.
XX
XX 21-SEP-2000; 2000US-0234274P.
XX
XX 25-SEP-2000; 2000US-0234997P.
XX
XX 25-SEP-2000; 2000US-0234998P.
XX
XX 26-SEP-2000; 2000US-0235484P.
XX
XX 27-SEP-2000; 2000US-0235834P.
XX
XX 27-SEP-2000; 2000US-0235836P.
XX
XX 29-SEP-2000; 2000US-0236327P.
XX
XX 29-SEP-2000; 2000US-0236367P.
XX

```

DR 29-SEP-2000; 2000US-0236368P.  
XX 29-SEP-2000; 2000US-0236369P.  
PT 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256717P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-02559678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-493426/52.  
XX DR

DR N-PSDB; AAK55435.  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX Claim 11; SEQ ID NO 10247; 3071pp + Sequence Listing; English.  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 32 AA;  
Query Match 95.0%; Score 19; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 VAEF 5  
DB 13 VAEF 16  
RESULT 32  
AAO9024  
ID AAO9024 standard; protein; 33 AA.  
AC AAO9024;  
XX 06-NOV-2001 (first entry)  
XX Human polypeptide SEQ ID NO 22916.  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorders; arthritis; inflammation.  
OS Homo sapiens.  
XX WO200164835-A2.  
XX 07-SEP-2001.  
XX 26-FEB-2001; 2001WO-US004927.  
XX 28-FEB-2000; 2000US-00515126.  
XX 18-MAY-2000; 2000US-00577409.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Drmanac RT;  
XX WPI; 2001-514838/56.  
XX N-PSDB; AAI86955.  
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
PT and treating e.g. leukemia, inflammation and immune disorders.  
XX Claim 20; SEQ ID NO 22916; 1399pp + Sequence Listing; English.  
XX PS

X The invention relates to human polynucleotides (AA179941-AA193841) and  
 C the encoded proteins (AA00010-AA013910) that exhibit activity relating to  
 C cytokine, cell proliferation or cell differentiation or which may induce  
 C production of other cytokines in other cell populations. The  
 C polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 C peptide therapy. The polypeptides have various cytokine-like activities,  
 C e.g. stem cell growth factor activity, haematopoiesis regulating  
 C activity, tissue growth factor activity, immunomodulatory activity and  
 C activin/inhibin activity and may be useful in the diagnosis and/or  
 C treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 C inflammation. Note: The sequence data for this patent did not form part  
 C of the printed specification, but was obtained in electronic format  
 C directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 33 AA;

Query Match 95.0%; Score 19; DB 4; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 2 VAEF 5  
 ||||  
 3 VAEF 6

# RESULT 33

AA77953  
 ID AA77953 standard; peptide; 35 AA.

AA77953;

09-OCT-1996 (first entry)

Antigenic Tbp2 peptide TBP2-21.

Tbp1; Tbp2; transferrin receptor operon; vaccine; antigen;  
 non-typable strain; Haemophilus influenzae; meningitis.

Synthetic.

WO9513370-A1.

18-MAY-1995.

07-NOV-1994; 94WO-CA000616.

08-NOV-1993; 93US-00148968.

29-DEC-1993; 93US-00175116.

(CONN-) CONNAUGHT LAB LTD.

Loosmore S, Harkness R, Schryvers A, Chong P, Gray-Owen S;  
 Yang Y, Mardin A, Klein M;

WPI; 1995-194089/25.

Nucleic acids encoding Haemophilus transferrin receptor - used to develop  
 prods for detection and in diagnosis, prevention and treatment of  
 Haemophilus infection.

Example 16; Page 72; 231pp; English.

AA77933-969 are predicted antigenic peptides derived from conserved  
 regions of the Tbp2 protein from H. influenzae strains Egan, Minna, DL63  
 and non-typable strain PAK12085. The transferrin receptor (TfR) operon  
 consists of two genes (Tbp1 and Tbp2) arranged in tandem and which are  
 transcribed from a single promoter. H. influenzae TfR is iron- and/or  
 haemin-regulated and a putative fur-binding site has been identified  
 upstream of Tbp2. Antibodies blocking this binding site may prevent  
 bacterial growth. Fragments of the TfR (or its genes) are useful in  
 vaccines to provide protection against, e.g. bacterial meningitis. An  
 advantage of using the TfR is that it shares homology with TfR of other

CC H. influenzae strains including non-typable strains. According to the  
 CC specification the present sequence shows residues 449-484 of Tbp2 from  
 CC the H. influenzae strain Egan

SQ Sequence 35 AA;

Query Match 95.0%; Score 19; DB 2; Length 35;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 14 VAEF 17

# RESULT 34

AAW46146

ID AAW46146 standard; protein; 35 AA.

AAW46146;

05-MAY-1998 (first entry)

Predicted antigenic Tbp1 peptide TBP2-21.

Transferrin receptor; Haemophilus influenzae type b; iron;  
 human transferrin; iron source; antibody; bacterial growth; vaccine;  
 immunogenic truncated analogue; antigen; Tbp1; Tbp2.

Synthetic.

Haemophilus influenzae.

WO9640929-A2.

19-DEC-1996.

07-JUN-1996; 96WO-CA000399.

07-JUN-1995; 95US-00483577.

17-MAY-1996; 96US-00649518.

(CONN-) CONNAUGHT LAB LTD.

Loosmore SM, Harkness RE, Schryvers AB, Chong P, Gray-Owen S;  
 Yang Y, Mardin AD, Klein MH;

WPI; 1997-052329/05.

Haemophilus truncated transferrin receptor protein analogue, Tbp2 - used  
 to induce protection against disease caused by transferrin producing  
 pathogens, or as antigen to detect Haemophilus TfR antibodies.

Example 16; Page 70; 228pp; English.

AAW46126-62 are predicted antigenic peptides derived from the Tbp2  
 protein of Haemophilus influenzae type b. Tbp2 is part of the transferrin  
 receptor, of which Tbp1 is also a subunit. The deduced amino acid  
 sequences of Tbp1 and Tbp2 were compared, and regions of conservation  
 identified. The above peptides are derived from these regions, the  
 present peptide being derived from residues 449-484. Iron is an essential  
 nutrient for the growth of these bacteria, and they can utilise human  
 transferrin as a source of iron. Antibodies which block the access of the  
 transferrin receptor to its iron source prevent bacterial growth. The  
 transferrin receptor, or fragments, therefore, are good vaccine  
 candidates. An immunogenic composition comprising (or encoding) the  
 immunogenic truncated analogue can be used to induce protection against a  
 disease caused by a bacterial pathogen that produces the transferrin  
 receptor. The immunogenic truncated analogue is also useful as an antigen  
 in immunoassays for the detection of Haemophilus transferrin receptor  
 antibodies, while the nucleic acid molecule can be used as a  
 hybridisation probe for the detection of other transferrin receptor genes  
 Sequence 35 AA;

Query Match 95.0%; Score 19; DB 2; Length 35;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 14 VAEF 17

RESULT 35  
 AAW51751  
 ID AAW51751 standard; protein; 35 AA.

AC AAW51751;  
 XX  
 XX  
 DT 13-JUN-2000 (first entry)  
 DE  
 DE H. Influenzae antigenic Tbp2 peptide TB22-21.  
 XX  
 KW Transferrin receptor; Tbp1; Tbp2; immunogenic; antibacterial; vaccine;  
 KW diagnosis.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 XX US6015688-A.  
 XX  
 XX 18-JAN-2000.  
 XX  
 XX 07-JUN-1995; 95US-00483577.  
 XX  
 XX 08-NOV-1993; 93US-00148968.  
 XX 29-DEC-1993; 93US-00175116.  
 XX 08-NOV-1994; 94US-00337483.  
 XX  
 XX (CONN-) CONNAUGHT LAB LTD.  
 XX  
 XX Loomore S, Harkness R, Schryvers A, Gray-Owen S, Yang Y;  
 XX Mardin A, Klein M, Chong P;  
 XX WPI; 1997-052329/05.

XX  
 XX Haemophilus truncated transferrin receptor protein analogue, Tbp2 - used  
 XX to induce protection against disease caused by transferrin producing  
 XX pathogens, or as antigen to detect Haemophilus rfr antibodies.  
 XX  
 XX Example 16; Col 39-40; 281pp; English.  
 XX  
 XX This invention describes a novel isolated and purified nucleic acid (I)  
 XX encoding an immunogenic, C-terminally truncated analog of one of the  
 XX transferrin receptor proteins Tbp1 or Tbp2 of Haemophilus influenzae  
 XX which has antibacterial activity. (I) are used for recombinant production  
 XX of truncated Tbp; as probes and primers for detecting, and diagnosing  
 XX infection by, Haemophilus, also for isolating similar sequences from  
 XX other bacteria; as immunogens for vaccinating against infections caused  
 XX by bacteria that produce transferrin receptors, e.g. Haemophilus,  
 XX Neisseria or Brachamella. The truncated proteins are useful as immunogens  
 XX (as above); for diagnosing infection (as antigens in immunoassays) and  
 XX for raising antibodies, used for diagnosis of infections or for passive  
 XX immunization. AAY51695-Y51767 represent H. influenzae transferrin  
 XX receptor proteins Tbp1 and Tbp2 antigenic peptide fragments  
 XX  
 XX Sequence 35 AA;

Query Match 95.0%; Score 19; DB 2; Length 35;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 14 VAEF 17

RESULT 37  
 AAW54105  
 ID AAW54105 standard; peptide; 35 AA.

AC AAW54105;  
 XX  
 XX  
 DT 20-JUL-1998 (first entry)  
 XX  
 DE Tbp2 antigenic peptide TBP2-21.

RESULT 36  
 AAW55788  
 ID AAW55788 standard; peptide; 35 AA.

AC AAW55788;  
 XX  
 XX  
 DT 14-JUL-1998 (first entry)  
 DE  
 DE Human aquaporin-1 NH2-terminal amino acid sequence CHIP28.  
 XX  
 KW Human; aquaporin-1; AQP-1; water channel protein; regulation;  
 KW osmotic change; erythrocyte; dryness; blindness; hydration; asthma;  
 KW secretion.  
 XX  
 OS Homo sapiens.

XX  
 XX Key Location/Qualifiers  
 XX FT Misc-difference 30 /label= Unknown  
 XX FT /note= "not specified but is given as Ser in the full  
 XX FT length protein given in AAW55786"  
 XX  
 XX US5741671-A.  
 XX  
 XX 21-APR-1998.  
 XX  
 XX 06-JUN-1995; 95US-00468763.  
 XX  
 XX 17-AUG-1992; 92US-00930168.  
 XX 24-FEB-1995; 95US-00339396.  
 XX  
 XX (UYJO ) UNIV JOHNS HOPKINS.  
 XX  
 XX Agre PC;  
 XX  
 XX WPI; 1998-260501/23.

XX  
 XX Polynucleotide(s) encoding water channel protein Aquaporin-1 - are useful  
 XX for recombinant production of protein for activity studies.  
 XX  
 XX Example 2; Col 20; 48pp; English.  
 XX  
 XX The present sequence represents the NH2-terminal amino acid sequence of  
 XX aquaporin-1 (AQP-1), isolated from human erythrocytes. AQP-1 is a water  
 XX channel protein. Water channel proteins regulate the passage of water in  
 XX and out of cells, in response to osmotic changes. The DNA encoding AQP-1  
 XX is useful for the recombinant production of AQP-1, found in mammalian  
 XX erythrocytes, and is useful in the study to identify reagents which  
 XX enhance or inhibit water channel function. This can lead to therapeutics  
 XX which enhance secretion e.g. in the case of dryness of eyes which can  
 XX lead to blindness or to hydrate large respiratory airways, as their  
 XX dryness can precipitate asthma  
 XX  
 XX Sequence 35 AA;

Query Match 95.0%; Score 19; DB 2; Length 35;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 14 VAEF 17

```

CX  tbp1; tbp2; vaccine; H. influenzae; antibody; diagnosis;
CW  passive immunisation; transferrin receptor operon.
CX  Haemophilus influenzae.
CX  US5708149-A.
CX  13-JAN-1998.
CX  07-JUN-1995; 95US-00487890.
CX  08-NOV-1993; 93US-00148968.
CX  29-DEC-1993; 93US-00175116.
CX  08-NOV-1994; 94US-00337483.
CX  (CONN-) CONNAUGHT LAB LTD.
CX  Gray-Owen S, Klein M, Harkness R, Loosmore S, Yang Y, Chong P;
XI  Murdin A, Schryvers A;
CX  WPI; 1998-100410/09.
CX  Purification of recombinant Haemophilus transferrin-binding protein - by
PT  solubilising inclusion bodies separated from cell lysate.
CX  Example 16/17; Column 37-38; 261pp; English.
CX  Peptides AAW53086-W53100 and peptides AAW54100-W54122 are derived from
CC  the Tbp2 protein. The Tbp2 protein is one of two proteins with genes
CC  found on the transferrin operon. These peptides can be used along with
CC  the genes, DNA sequences and recombinant proteins for diagnosis,
CC  immunisation and the generation of diagnostic and immunological reagents.
CC  They can also be used to protect from bacteria that produce transferrin
CC  receptor protein
CX  Sequence 35 AA;
CX  Query Match 95.0%; Score 19; DB 2; Length 35;
CX  Best Local Similarity 100.0%; Pred. No. 4.7e+02;
CX  Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CX  2 VAEF 5
CX  14 VAEF 17
CX  RESULT 38
CX  ID AAW94320 standard; peptide; 35 AA.
CX  AC AAW94320;
CX  DT 12-APR-1999 (first entry)
CX  DE Human aquaporin-1 (CHIP28) N-terminal peptide.
CX  KW Rat; aquaporin-5; AQP5; AQP1; transmembrane water channel protein;
KW  major intrinsic protein; MIP; CHIP28.
CX  OS Homo sapiens.
CX  FH Key Location/Qualifiers
CX  FT Misc-difference 30
CX  FT /note= "unspecified"
CX  PN US5858702-A.
CX  PD 12-JAN-1999.
CX  PF 24-FEB-1995; 95US-00393996.
CX  XX 13-DEC-1991; 91US-00806273.
CX  PR
CX  PR 17-AUG-1992; 92US-00930168.
CX  PA (UYJO ) UNIV JOHNS HOPKINS.
CX  PI Agre PC;
CX  XX WPI; 1999-152100/13.
CX  DR DNA encoding aquaporin-5 water channel protein - useful for producing
XX  recombinant protein.
CX  PS Example 2; Col 20; 47pp; English.
CX  XX The present invention describes a water channel protein called aquaporin-
CC  5 (AQP5). The polynucleotide encoding AQP5 is useful for producing
CC  recombinant AQP5, which can be incorporated into proteoliposomes or cell
CC  membrane vesicles, which are able to be used in screening assays for water
CC  channel agonists or antagonists. The present sequence represents the N-
CC  terminal peptide from AQP1 (also called CHIP28), from an example of the
CC  present invention
CX  XX Sequence 35 AA;
CX  Query Match 95.0%; Score 19; DB 2; Length 35;
CX  Best Local Similarity 100.0%; Pred. No. 4.7e+02;
CX  Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CX  2 VAEF 5
CX  14 VAEF 17
CX  RESULT 39
CX  ID AAY80448 standard; peptide; 35 AA.
CX  AC AAY80448;
CX  DT 06-JUN-2000 (first entry)
CX  DE H. influenzae transferrin receptor Tbp1 epitope TBP2-21.
CX  KW Antibacterial; antiinflammatory; auditory; respiratory; antibody;
KW  antiserum; transferrin receptor; immunogen; epitope; otitis media;
KW  bacterial meningitis; epiglottitis; pneumonia; tracheobronchitis.
CX  OS Haemophilus influenzae.
CX  XX US6008326-A.
CX  PN 28-DEC-1999.
CX  XX 07-JUN-1995; 95US-00474671.
CX  XX 08-NOV-1993; 93US-00148968.
CX  PR 29-DEC-1993; 93US-00175116.
CX  PR 08-NOV-1995; 95US-00337483.
CX  XX (CONN-) CONNAUGHT LAB LTD.
CX  XX Loosmore S, Harkness R, Chong P, Gray-Owen S, Yang Y, Klein M;
PI  Murdin A, Schryvers A;
CX  XX WPI; 2000-096387/08.
CX  XX Antibodies specific for transferrin receptor proteins of Haemophilus
PT  influenzae, useful for treating otitis media, epiglottitis, pneumonia and
PT  tracheobronchitis.
CX  PS Disclosure; Col 37-38; 252pp; English.
CX  XX The invention relates to novel antibodies (or monospecific antisera)
CC  specific for single transferrin receptor proteins (or immunogenic
CC  specific for single transferrin receptor proteins (or immunogenic

```



PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229350P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 08-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 26-SEP-2000; 2000US-0234998P.  
PR 27-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0241826P.  
PR 08-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.

PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251858P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX WPI; 2001-465569/50.  
XX N-PSDB; AA198806.  
PT Isolated nucleic acid molecule encoding excretory system antigen is used  
PT in preventing, treating or ameliorating a medical condition.  
PS Claim 11; SEQ ID NO 570; 574pp + Sequence Listing; English.  
XX The invention relates to novel excretory system related human  
CC polynucleotides (AA198567-AA199503) and the encoded proteins (AA199594-  
CC AA199913) useful for preventing, treating or ameliorating medical  
CC conditions e.g. by protein or gene therapy, especially disorders related  
CC to the excretory system. The genes are isolated from a range of human  
CC tissues disclosed in the specification. The nucleic acids, proteins,  
CC antibodies and (ant)agonists are useful in the diagnosis, treatment and  
CC prevention of: (a) cancer, e.g. breast and ovarian cancer and other  
CC cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal  
CC tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's  
CC disease, allergies, autoimmune haemolytic anaemia, autoimmune  
CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,  
CC rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders  
CC such as myocardial ischaemia; (d) wound healing; (e) neurological  
CC diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases  
CC such as viral, bacterial, fungal and parasitic infections. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 37 AA;

Query Match 95.0%; Score 19; DB 4; Length 37;

Best Local Similarity 100.0%; Pred. No. Se+0; 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5

Db 14 VAEF 17  
RESULT 42  
AAW42648  
ID AAW42648 standard; protein; 37 AA.  
XX AC AAW42648;  
XX DT 22-OCT-2001 (first entry)  
XX DE Human kidney related polypeptide SEQ ID NO 517.  
XX KW Human; kidney antigen; immunosuppressive; antiarthritic; antirheumatic;  
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;  
KW ophthalmological; anti-allergic; hepatotropic; antidiabetic;  
KW anti-inflammatory; antitumor; anticonvulsant; antiparasitic;  
KW gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection.  
XX OS Homo sapiens.  
XX DN WO200155323-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US001343.  
XX PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180828P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220363P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 03-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0244674P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246521P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.



```

XX PF 30-JAN-2001; 2001WO-US0000662.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-496933/54.
XX XX
XX PT New spatially-addressable set of single exon nucleic acid probes, useful
XX PT for measuring gene expression in sample derived from human breast,
XX PT comprises number of single exon nucleic acid probes.
XX PS Claim 27; SEQ ID NO 14330; 327pp + Sequence Listing; English.
XX CC The invention relates to a spatially-addressable set of single exon
XX CC nucleic acid probes for measuring gene expression in a sample derived
XX CC from human breast and BT 474 cells. The method involves contacting the
XX CC probes with a collection of detectably labelled nucleic acids derived
XX CC from mRNA of human breast, and then measuring the label bound to each
XX CC probe of the microarray. The probes are useful for verifying the
XX CC expression of regions of genomic DNA predicted to encode proteins. They
XX CC are useful for gene discovery, and for determining predisposition and/or
XX CC prognosing breast disease. Gene expression analysis is useful for
XX CC assessing the toxicity of chemical agents on cells. The microarray of
XX CC this invention presents a far greater diversity of probes for measuring
XX CC gene expression, with far less bias than expressed sequence tag
XX CC microarrays. The method is suitable for rapid production of functional
XX CC information from genomic sequence. The present sequence is a peptide
XX CC encoded by a single exon nucleic acid probe of the invention. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 43 AA;

Query Match 95.0%; Score 19; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 5.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5
Db 40 VAEF 43

RESULT 45
ABB21904
ID ABB21904 standard; protein; 43 AA.
XX AC ABB21904;
XX XX
XX DT 23-JAN-2002 (first entry)
XX DE Protein #3903 encoded by probe for measuring heart cell gene expression.
XX KW Human; gene expression; heart; microarray; vascular system;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease.
XX OS Homo sapiens.
XX XX WO200157274-A2.
XX PN 09-AUG-2001.
XX PD
XX PF 30-JAN-2001; 2001WO-US0000664.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-488999/53.
XX XX
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT hearts.
XX PS Claim 15; SEQ ID NO 23674; 530pp; English.
XX CC The present invention relates to single exon nucleic acid probes for
XX CC measuring human gene expression in a sample derived from human heart (see
XX CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX CC probe. The probes may be used for predicting, measuring and displaying
XX CC gene expression in samples derived from the human heart via microarrays.
XX CC By measuring gene expression, the probes are useful for predicting,
XX CC diagnosing, grading, staging, monitoring and prognosing diseases of the
XX CC human heart and vascular system e.g. cardiovascular disease,
XX CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 43 AA;

Query Match 95.0%; Score 19; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 5.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5
Db 40 VAEF 43

RESULT 46
ABG51414
ID ABG51414 standard; peptide; 43 AA.
XX AC ABG51414;
XX XX
XX DT 25-FEB-2003 (first entry)
XX DE Human liver peptide, SEQ ID NO 30062.
XX XX
XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX KW hypercholesterolaemia; coronary heart disease.
XX OS Homo sapiens.
XX XX WO200157273-A2.
XX PN 09-AUG-2001.
XX PD
XX PF 30-JAN-2001; 2001WO-US0000664.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.

```

(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI; 2001-488898/53.  
Human genome-derived single exon nucleic acid probes useful for analyzing  
gene expression in human adult liver.  
Claim 27; SEQ ID NO 30662; 658pp; English.  
The invention relates to a single exon nucleic acid probe (SENP) (I) for  
measuring human gene expression in a sample derived from human adult  
liver, comprising one of 13109 defined nucleotide sequences given in the  
specification (or complements/ fragments). The probe hybridises at high  
stringency to a nucleic acid molecule expressed in the human adult liver.  
(I) may be used for predicting, measuring and displaying gene expression  
in samples derived from human adult liver. The genes identified may be  
involved in genetic liver diseases such as cirrhosis,  
hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
associated with coronary heart disease. ABQ47348-ABQ59930 represent human  
liver single exon encoded peptides of the invention. Note: The sequence  
information for this patent does not appear in the printed specification  
but was obtained in electronic format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences  
Sequence 43 AA;  
Query Match 95.0%; Score 19; DB 4; Length 43;  
Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0  
NY 2 VAEF 5  
|||  
b 40 VAEF 43  
RESULT 47  
AM05214  
D AM05214 standard; protein; 43 AA.  
X AM05214;  
X  
Y 09-OCT-2001 (first entry)  
X  
X Peptide #3896 encoded by probe for measuring breast gene expression.  
X Probe; human; breast disease; breast cancer; development disorder;  
X inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
X Homo sapiens.  
X WO200157270-A2.  
X  
X 09-AUG-2001.  
X  
X 29-JAN-2001; 2001WC-US000661.  
X  
X 04-FEB-2000; 2000US-0180312P.  
X 26-MAY-2000; 2000US-0207456P.  
X 30-JUN-2000; 2000US-00608408.  
X 03-AUG-2000; 2000US-00632366.  
X 21-SEP-2000; 2000US-0234687P.  
X 27-SEP-2000; 2000US-0236359P.  
X 04-OCT-2000; 2000GB-00024263.  
X  
X (MOLE-) MOLECULAR DYNAMICS INC.  
X Penn SG, Hanzel DK, Chen W, Rank DR;  
X WPI; 2001-476286/51.  
X

PT Novel single exon nucleic acid probe used to measuring gene expression in  
PT a human breast.  
XX  
XX Claim 27; SEQ ID NO 13954; 322pp; English.  
XX  
CC The present invention relates to novel single exon nucleic acid probes  
CC (see AA100010-AA110067). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for measuring human gene expression in a  
CC human breast sample, where the probe hybridises at high stringency to a  
CC nucleic acid expressed in the human breast. The probes are useful for  
CC predicting, diagnosing, grading, staging, monitoring and prognosing  
CC diseases of the human breast, particularly those diseases with polygenic  
CC aetiology. The diseases include: breast cancer, disorders of development,  
CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
CC breast disease and non-carcinoma tumours. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 43 AA;  
XX

Query Match 95.0%; Score 19; DB 4; Length 43;  
Best Local Similarity 100.0%; Pred. No. 5.9e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0;

QY 2 VAEF 5  
DB 40 VAEF 43  
|||||

RESULT 48  
AAB27178  
ID AAB27178 standard; protein; 47 AA.  
AC AAB27178;  
XX  
DT 27-FEB-2001 (first entry)  
XX  
DE RSV partial protein sequence SEQ ID NO: 26.  
XX  
XX Negative stranded RNA virus; vaccine; attenuated virus; RSV; PIV;  
KW measles; respiratory syncytial virus; parainfluenza virus.  
XX  
XX Respiratory syncytial virus.  
OS  
XX  
XX WC2000061737-A2.  
PN  
XX  
XX 19-OCT-2000.  
PD  
XX  
XX 12-APR-2000; 2000WO-US009695.  
PF  
XX  
XX 13-APR-1999; 99US-0129006P.  
PR  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA  
XX  
XX Murphy BR, Collins PL, Durbin AP, Skiadopoulos MH;  
PI  
XX  
XX WPI; 2000-687044/67.  
DR  
XX  
XX Producing attenuated negative stranded RNA virus vaccines from cloned  
PT sequences, useful for immunizing against e.g. respiratory syncytial  
PT virus, human parainfluenza virus, Sendai virus Newcastle disease virus,  
PT mumps virus and measles virus.  
XX  
XX Example 1; Page 62; 137pp; English.  
XX  
XX The present invention is concerned with producing vaccines against  
CC negative stranded RNA viruses. These viruses include measles, respiratory  
CC syncytial virus (RSV) and parainfluenza virus (PIV) in particular. The  
CC method of the invention comprises the production of a mutated form of the  
CC virus which attenuates the strain and enables it to be used as a vaccine.  
CC The present sequence comprises a partial viral protein sequence  
XX

```

SQ Sequence 47 AA;
  Query Match          95.0%; Score 19; DB 3; Length 47;
  Best Local Similarity 100.0%; Pred. No. 6.4e+02;
  Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 VAEF 5
      |||||
Db      34 VAEF 37

RESULT 49
AAM18541
ID AAM18541 standard; protein; 48 AA.
XX
AC AAM18541;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #4975 encoded by probe for measuring cervical gene expression.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000670.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
  gene expression in human cervical epithelial cells.
XX
PS Claim 27; SEQ ID NO 23367; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
  (SENPs: see AAM10068-AAM28459). The present sequence is a peptide encoded
  by one such probe. The SENPs are derived from human HeLa cells. The SENPs
  can be used to produce a single exon microarray, which can be used for
  measuring human gene expression in a sample derived from human cervical
  epithelial cells. By measuring gene expression, the probes are therefore
  useful in grading and/or staging of diseases of the cervix, notably
  cervical cancer. Note: The sequence data for this patent did not form
  part of the printed specification, but was obtained in electronic format
  directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 48 AA;
  Query Match          95.0%; Score 19; DB 4; Length 48;
  Best Local Similarity 100.0%; Pred. No. 6.6e+02;
  Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 VAEF 5
      |||||
Db      3 VAEF 6

RESULT 50
ABB37578
ID ABB37578 standard; peptide; 48 AA.
XX
AC ABB37578;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #5084 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
  gene expression in human fetal liver.
XX
PS Claim 27; SEQ ID NO 30213; 639pp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
  human gene expression in a sample derived from human foetal liver. The
  single exon nucleic acid probes may be used for predicting, measuring and
  displaying gene expression in samples derived from human fetal liver. The
  present sequence is a peptide encoded by a single exon nucleic acid probe
  of the invention. Note: The sequence data for this patent did not form
  part of the printed specification, but was obtained in electronic format
  directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 48 AA;
  Query Match          95.0%; Score 19; DB 4; Length 48;
  Best Local Similarity 100.0%; Pred. No. 6.6e+02;
  Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 VAEF 5
      |||||
Db      3 VAEF 6

RESULT 51
AAM30996
ID AAM30996 standard; protein; 48 AA.
XX
AC AAM30996;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #5033 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.

```

X WO200157272-A2.  
 N 09-AUG-2001.  
 X 30-JAN-2001; 2001WO-US000663.  
 F 04-FEB-2000; 2000US-0180312P.  
 X 26-MAY-2000; 2000US-0207456P.  
 R 30-JUN-2000; 2000US-00608408.  
 R 03-AUG-2000; 2000US-00632366.  
 R 21-SEP-2000; 2000US-0234587P.  
 R 27-SEP-2000; 2000US-0236359P.  
 R 04-OCT-2000; 2000GB-00024263.  
 X (MOLE-) MOLECULAR DYNAMICS INC.  
 A Penn SG, Hanzel DK, Chen W, Rank DR;  
 X WPI; 2001-488997/53.  
 X Human genome-derived single exon nucleic acid probes useful for analyzing  
 T gene expression in human placenta.  
 T Claim 27; SEQ ID NO 31265; 654pp; English.  
 S The present invention relates to single exon nucleic acid probes (SENP:  
 C see AA131315-AA157546). The present sequence is a peptide encoded by one  
 C such probe. The probes are useful for producing a microarray for  
 C predicting, measuring and displaying gene expression in samples derived  
 C from human placenta. The probes are useful for antenatal diagnosis of  
 C human genetic disorders  
 X Sequence 48 AA;  
 Q Query Match 95.0%; Score 19; DB 4; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 2 VAEF 5  
 b |||||  
 3 VAEF 6  
 RESULT 52  
 LBB22868  
 D ABB22868 standard; protein; 48 AA.  
 X ABB22868;  
 C 23-JAN-2002 (first entry)  
 Y Protein #4867 encoded by probe for measuring heart cell gene expression.  
 X Human; gene expression; heart; microarray; vascular system;  
 W cardiovascular disease; hypertension; cardiac arrhythmia;  
 W congenital heart disease.  
 X Homo sapiens.  
 X WO200157274-A2.  
 N 09-AUG-2001.  
 X 30-JAN-2001; 2001WO-US000666.  
 F 04-FEB-2000; 2000US-0180312P.  
 X 26-MAY-2000; 2000US-0207456P.  
 R 30-JUN-2000; 2000US-00608408.  
 R 03-AUG-2000; 2000US-00632366.  
 R 21-SEP-2000; 2000US-0234587P.  
 R 27-SEP-2000; 2000US-0236359P.  
 R 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488997/53.  
 XX Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts.  
 PT Claim 15; SEQ ID NO 24638; 530pp; English.  
 XX The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart (see  
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying  
 CC gene expression in samples derived from the human heart via microarrays.  
 CC By measuring gene expression, the probes are useful for predicting,  
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 CC human heart and vascular system e.g. cardiovascular disease,  
 CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 48 AA;  
 SQ Query Match 95.0%; Score 19; DB 4; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 Db |||||  
 3 VAEF 6  
 RESULT 53  
 AAM70687  
 ID AAM70687 standard; protein; 48 AA.  
 X AAM70687;  
 AC 06-NOV-2001 (first entry)  
 X Human bone marrow expressed probe encoded protein SEQ ID NO: 30993.  
 DE Human; bone marrow expressed exon; gene expression analysis; probe;  
 XW microarray; cancer; leukaemia; lymphoma; myeloma.  
 XW Homo sapiens.  
 OS WO200157276-A2.  
 XN 09-AUG-2001.  
 X 30-JAN-2001; 2001WO-US000668.  
 XX 04-FEB-2000; 2000US-0180312P.  
 XX 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234587P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488997/53.  
 XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human bone marrow.

```
XX PS Example 4; SEQ ID NO 30993; 658pp + Sequence Listing; English.
XX CC
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention
XX SQ Sequence 48 AA;
XX
XX Query Match 95.0%; Score 19; DB 4; Length 48;
XX Best Local Similarity 100.0%; Pred. No. 6.6e+02;
XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 VAEF 5
XX Db |||||
XX 3 VAEF 6
XX
XX RESULT 54
XX ABG52389
XX ID ABG52389 standard; peptide; 48 AA.
XX AC ABG52389;
XX
XX DT 25-FEB-2003 (first entry)
XX
XX DE Human liver peptide, SEQ ID NO 31037.
XX
XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX KW hypercholesterolaemia; coronary heart disease.
XX
XX OS Homo sapiens.
XX
XX PN WO200157273-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000664.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-488898/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human adult liver.
XX
XX PS Claim 27; SEQ ID NO 31037; 658pp; English.
XX
XX CC The invention relates to a single exon nucleic acid probe (SEN) (I) for
XX CC measuring human gene expression in a sample derived from human adult
XX CC liver, comprising one of 13109 defined nucleotide sequences given in the
XX CC specification (or complements/ fragments). The probe hybridises at high
XX CC stringency to a nucleic acid molecule expressed in the human adult liver.
XX CC (I) may be used for predicting, measuring and displaying gene expression
XX CC in samples derived from human adult liver. The genes identified may be
XX CC involved in genetic liver diseases such as cirrhosis,
XX CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX CC associated with coronary heart disease. ABG47348-ABG5930 represent human
XX CC liver single exon encoded peptides of the invention. Note: The sequence
XX CC information for this patent does not appear in the printed specification
```

```
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 48 AA;
XX
XX Query Match 95.0%; Score 19; DB 4; Length 48;
XX Best Local Similarity 100.0%; Pred. No. 6.6e+02;
XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 VAEF 5
XX Db |||||
XX 3 VAEF 6
XX
XX RESULT 55
XX AAM06107
XX ID AAM06107 standard; protein; 48 AA.
XX
XX AC AAM06107;
XX
XX DT 09-OCT-2001 (first entry)
XX
XX DE Peptide #4789 encoded by probe for measuring breast gene expression.
XX
XX KW Probe; human; breast disease; breast cancer; development disorder;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX OS Homo sapiens.
XX
XX PN WO200157270-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 29-JAN-2001; 2001WO-US000661.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-476286/51.
XX
XX PT Novel single exon nucleic acid probe used to measuring gene expression in
XX PT a human breast.
XX
XX PS Claim 27; SEQ ID NO 14847; 322pp; English.
XX
XX CC The present invention relates to novel single exon nucleic acid probes
XX CC (see AAI00010-AAI10067). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for measuring human gene expression in
XX CC a human breast sample, where the probe hybridises at high stringency to a
XX CC nucleic acid expressed in the human breast. The probes are useful for
XX CC predicting diagnosing, grading, staging, monitoring and prognosing
XX CC diseases of the human breast, particularly those diseases with polygenic
XX CC aetiology. The diseases include: breast cancer; disorders of development,
XX CC inflammatory diseases of the breast, fibrocystic changes, proliferative,
XX CC breast disease and non-carcinoma tumours. Note: The sequence data for
XX CC this patent did not form part of the printed specification, but was
XX CC obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 48 AA;
XX
XX Query Match 95.0%; Score 19; DB 4; Length 48;
XX Best Local Similarity 100.0%; Pred. No. 6.6e+02;
XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

AC	AAO05247;	
XX	06-NOV-2001 (first entry)	
DT		
XX	Human polypeptide SEQ ID NO 19139.	
DE		
XX	Human; cytokine; cell proliferation; cell differentiation; gene therapy;	
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;	
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;	
KW	nervous system disorders; arthritis; inflammation.	
XX	Homo sapiens.	
OS		
XX	WO200164835-A2.	
PN		
XX	07-SEP-2001.	
PD		
XX	26-FEB-2001; 2001WO-US004927.	
PF		
XX	28-FEB-2000; 2000US-00515126.	
XX	18-MAY-2000; 2000US-00577409.	
PR		
XX	(HYSE-) HYSEQ INC.	
PA		
XX	Tang YT, Liu C, Drmanac RT;	
PI		
XX	WPI; 2001-514838/56.	
DR	N-PSDB; AAI85178.	
DR		
XX	Isolated nucleic acids and polypeptides, useful for preventing diagnosing	
PT	and treating e.g. leukemia, inflammation and immune disorders.	
PT		
XX	Claim 20; SEQ ID NO 19139; 1399pp + Sequence Listing; English.	
PS		
XX	The invention relates to human polynucleotides (AAI79941-AAI93841) and	
CC	the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to	
CC	cytokine, cell proliferation or cell differentiation or which may induce	
CC	production of other cytokines in other cell populations. The	
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or	
CC	peptide therapy. The polypeptides have various cytokine-like activities,	
CC	e.g. stem cell growth factor activity, haematopoiesis regulating	
CC	activity, tissue growth factor activity, immunomodulatory activity and	
CC	activin/inhibin activity and may be useful in the diagnosis and/or	
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and	
CC	inflammation. Note: The sequence data for this patent did not form part	
CC	of the printed specification, but was obtained in electronic format	
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	
XX		
SQ	Sequence 53 AA;	
	Query Match 95.0%; Score 19; DB 4; Length 53;	
	Best Local Similarity 100.0%; Pred. No. 7.3e+02;	
	Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
QY	2 VAEF 5	
Db	5 VAEF 8	
RESULT 58		
AAI87390		
ID	AAI87390 standard; protein; 56 AA.	
XX		
AC	AAI87390;	
XX		
DT	07-NOV-2001 (first entry)	
XX	Human immune/haematopoietic antigen SEQ ID NO:14983.	
DE		
XX	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	
KW	cytostatic; gene therapy; vaccine; metastasis.	
KW		
XX	Homo sapiens.	
OS		



I Rosen CA, Barash SC, Ruben SM;  
 K WPI; 2001-483426/52.  
 R N-PSDB; AAK60171.  
 K Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 K useful for preventing, diagnosing and/or treating cancers and metastasis.  
 L Claim 11; SEQ ID NO 14983; 3071pp + Sequence Listing; English.  
 S  
 X AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 C amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 C activity, and can be used in gene therapy and vaccine production. (I)  
 C proteins and polynucleotides may be used in the prevention, diagnosis and  
 C treatment of diseases associated with inappropriate (I) expression. For  
 C example, they may be used to treat disorders associated with decreased  
 C expression by rectifying mutations or deletions in a patient's genome  
 C that affect the activity of (I) by expressing inactive proteins or to  
 C supplement the patients own production of (I). Additionally, (I)  
 C polynucleotides may be used to produce the secreted (I), by inserting the  
 C nucleic acids into a host cell and culturing the cell to prevent the  
 C protein. (I) proteins and polynucleotides may be used to prevent,  
 C diagnose and treat immune/hematopoietic-related diseases, especially  
 C cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
 C to AAK87694 represent human immune/hematopoietic antigen genomic  
 C sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 C represent sequences used in the exemplification of the present invention  
 X Sequence 56 AA;  
 Q

Query Match 95.0%; Score 19; DB 4; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5  
 b 1 VAEF 4

RESULT 59

BP97126  
 D ABP97126 standard; peptide; 56 AA.

C C

C ABP97126;

24-JUN-2003 (first entry)

Human matrix metalloproteinase 7 cleavage region peptide SEQ ID NO:4.  
 Human; matrix metalloproteinase; MMP; anticancer; wound healing;  
 matrix metalloproteinase inhibitor; antitumor; angiogenic; cardiac;  
 vascular endothelial growth factor inhibitor; VEGF inhibitor; cytostatic;  
 vulnary; cerebroprotective; antidiabetic; ophthalmological; tumor;  
 dermatological; metastatic; non-metastatic; vascularised; heart disease;  
 non-vascularised; surgical incision; chronic wound; stroke; angiogenesis;  
 macular degeneration; diabetic retinopathy; cleavage region.

Homo sapiens.

WO2003018748-A2.

06-MAR-2003.

15-AUG-2002; 2002WO-US026319.

16-AUG-2001; 2001US-0312726P.

21-DEC-2001; 2001US-00032376.

21-MAY-2002; 2002US-00153185.

(KIMB ) KIMBERLY-CLARK WORLDWIDE INC.

Quirk S, Weart IF;

WPI; 2003-381408/36.

XX Anti-angiogenic composition comprising peptide inhibitor of matrix  
 PT metalloproteinase, useful for decreasing the expression of vascular  
 PT endothelial growth factor and treating cancers and tissue injuries.  
 XX Claim 17; Page 15; 103pp; English.

XX The present invention describes an anti-angiogenic composition (I) for  
 CC inhibiting expression of vascular endothelial growth factor (VEGF). (I)  
 CC comprises an effective amount of a peptide inhibitor of matrix  
 CC metalloproteinase (MMP), where the peptide can inhibit the expression of  
 CC VEGF. (I) has cytostatic, vulnary, cardiac, cerebroprotective,  
 CC antidiabetic, ophthalmological and dermatological activities. (I) can be  
 CC used for inhibiting expression of VEGF, and so can be used for inhibiting  
 CC growth of tumors and diminishing tumours size. The tumour can be  
 CC metastatic, non-metastatic, vascularised, non-vascularised, hard or soft.  
 CC (I) is also useful for treating injuries including wounds, surgical  
 CC incisions, chronic wounds, heart diseases and stroke. (I) is also useful  
 CC for treating disorders characterised by excessive angiogenesis e.g.  
 CC macular degeneration and diabetic retinopathy. The present sequence  
 CC represents a human MMP cleavage region peptide, which is used in the  
 CC exemplification of the present invention

Sequence 56 AA;

Query Match 95.0%; Score 19; DB 6; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 Db 31 VAEF 34

RESULT 60

ABG76312

ID ABG76312 standard; protein; 56 AA.

XX AC ABG76312;

XX 10-MAY-2003 (first entry)

Human matrix metalloproteinase (MMP) peptide inhibitor #4.

Human; peptide inhibitor; matrix metalloproteinase-7; MMP-7;  
 cleavage region; proenzyme form; cellular proliferation; fibroblast;  
 keratinocyte; healthy skin development; wound healing; scarring;  
 skin tone; wrinkle; anti-aging; vulnary.

Homo sapiens.

WO2003016520-A1.

27-FEB-2003.

15-AUG-2002; 2002WO-US026198.

16-AUG-2001; 2001US-0312726P.

21-DEC-2001; 2001US-00032376.

21-MAY-2002; 2002US-00153185.

(KIMB ) KIMBERLY-CLARK WORLDWIDE INC.

Quirk S, Villanueva JM;

WPI; 2003-289980/28.

Novel peptide inhibitor of proteinase activity of matrix  
 metalloproteinases, e.g. matrix metalloproteinase-2, useful for  
 stimulating cellular proliferation of fibroblasts or keratinocytes.  
 Claim 1; Page 16; 120pp; English.

XX CC The present invention relates to peptide inhibitors of metalloproteinases  
 CC (MMPs), particularly metalloproteinase-2 (MMP-2). The inhibitors have  
 CC peptide sequences related to the cleavage regions of the proenzyme forms  
 CC of the MMPs. The peptide inhibitors are useful for stimulating cellular  
 CC proliferation of fibroblasts or keratinocytes, promoting healthy skin  
 CC development, treating wounds, preventing scarring, improving skin tone,  
 CC reducing wrinkling and for stimulating the development of smooth, healthy  
 CC skin. The peptide inhibitors are useful as anti-aging and wound healing  
 CC compounds. ABG76309-ABG76321 represent peptide inhibitors of MMPs  
 XX CC

SQ Sequence 56 AA;

Query Match 95.0%; Score 19; DB 6; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VAEF 5  
 DB 31 VAEF 34

RESULT 61

AAW55888

ID AAW55888 standard; protein; 58 AA.

XX AC AAW55888;

XX DT 22-JUL-1998 (first entry)

XX DE Rat PC12.

XX KW Rat; telomerase; human; cancer; screening; inhibitor; elucidation;

XX KW detection; probe; diagnosis; cell growth; ageing.

XX OS Rattus sp.

XX PN WO9807838-A1.

XX PD 26-FEB-1998.

XX PF 21-AUG-1997; 97WO-JP002904.

XX PR 21-AUG-1996; 96JP-00219761.

XX PR 31-JAN-1997; 97JP-00018878.

XX PR 17-FEB-1997; 97JP-00031807.

XX PA (MITU) MITSUBISHI CHEM CORP.

XX PI Ishikawa F, Nakamura H, Takahashi K, Fujino Y, Harada N;

XX DR WPI; 1998-169149/15.

XX DR N-PSDB; AAV25990.

XX PT Telomerase protein of higher animals and humans and gene encoding it -

XX PT for use in diagnosis of cancer, screening of telomerase inhibitors and

XX PT elucidation of biological control mechanisms.

XX PS Example 1; Page 69-70; 106pp; Japanese.

XX CC The present sequence represents rat PC12 which is used in an example of

XX CC the present invention which describes protein components of telomerase.

XX CC The DNA or RNA encoding the telomerase protein component or its fragments

XX CC can be used as a nucleotide probe for the detection of cancer cells and

XX CC for diagnosis of cancer. Potential telomerase inhibitors can be screened

XX CC by measuring their effect on the assay of the active form in cells or

XX CC tissues. The polypeptide and DNA coding for it can be used in the

XX CC elucidation of biological control mechanisms of, e.g. cell growth or

XX CC ageing and of the mechanisms of cancer development

XX CC

XX SQ Sequence 58 AA;

XX Query Match 95.0%; Score 19; DB 2; Length 58;

Best Local Similarity 100.0%; Pred. No. 8.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VAEF 5  
 DB 49 VAEF 52

RESULT 62

ABP04854

ID ABP04854 standard; protein; 58 AA.

XX AC ABP04854;

XX DT 24-JUN-2002 (first entry)

XX DE Human ORFX protein sequence SEQ ID NO:9690.

XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;

XX KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;

XX KW degenerative disorder; osteoarthritis; neurodegenerative disorder;

XX KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;

XX KW hypertension; hypothyroidism; cholesterol ester storage disease;

XX KW immune deficiency; immune disorder; infectious disease;

XX KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;

XX KW myasthenia gravis.

XX OS Homo sapiens.

XX PN WO200192523-A2.

XX PD 06-DEC-2001.

XX PF 29-MAY-2001; 2001WO-US010836.

XX PR 30-MAY-2000; 2000US-0206132P.

XX PR 29-AUG-2000; 2000US-0228716P.

XX PA (CURA-) CURAGEN CORP.

XX PI Shimkets RA, Leach MD;

XX DR WPI; 2002-106308/14.

XX DR N-PSDB; ABN20606.

XX PT Novel human polypeptides and polynucleotides useful for diagnosing,

XX PT preventing and treating cardiovascular disease, neurodegenerative,

XX PT hyperproliferative disorders and autoimmune disorders.

XX PS Disclosure; SEQ ID NO 9690; 1037pp; English.

XX CC The present invention describes substantially purified human proteins

XX CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1

XX CC in the specification). ABN15762 to ABN27252 encode the human ORFX

XX CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for

XX CC treating or preventing a pathology associated with an ORFX-associated

XX CC disorder in humans, and in the manufacture of a medicament for treating a

XX CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide

XX CC sequences can be used in gene therapy. ORFX sequences can be used in the

XX CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,

XX CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,

XX CC osteoarthritis, neurodegenerative disorders, diabetes mellitus, systemic

XX CC lupus erythematosus, cardiovascular diseases, hypothyroidism, cholesterol ester

XX CC storage disease, various immune deficiencies and disorders, infectious

XX CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid

XX CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host

XX CC disease and autoimmune inflammatory eye disease. ORFX proteins are also

XX CC useful for treating burns, incisions, ulcers, for treating osteoporosis,

XX CC bone degenerative disorders, or periodontal disease, and for gut

XX CC protection or regeneration and treatment of lung or liver fibrosis,

XX CC reperfusion injury in various tissues and conditions resulting from

XX CC systemic cytokine damage. N.B. The sequence data for this patent did not

C form part of the printed specification, but was obtained in electronic  
C format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
X Sequence 58 AA;  
Q Query Match 95.0%; Score 19; DB 5; Length 58;  
Best Local Similarity 100.0%; Pred. No. 8.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 2 VAEF 5  
b 30 VAEF 33  
RESULT 63  
AG60085  
D AAG60085 standard; protein; 59 AA.  
C AAG60085;  
X X  
T 18-OCT-2000 (first entry)  
E Arabidopsis thaliana protein fragment SEQ ID NO: 77787.  
X  
W Protein identification; signal transduction pathway; metabolic pathway;  
W hybridisation assay; genetic mapping; gene expression control; promoter;  
W termination sequence.  
X X  
S Arabidopsis thaliana.  
X  
N EP1033405-A2.  
X  
D 06-SEP-2000.  
X  
F 25-FEB-2000; 2000EP-00301439.  
X  
R 25-FEB-1999; 99US-0121825P.  
R 05-MAR-1999; 99US-0123180P.  
R 09-MAR-1999; 99US-0123548P.  
R 23-MAR-1999; 99US-0125788P.  
R 25-MAR-1999; 99US-0126264P.  
R 28-MAR-1999; 99US-0126785P.  
R 01-APR-1999; 99US-0127462P.  
R 06-APR-1999; 99US-0128234P.  
R 08-APR-1999; 99US-0128714P.  
R 16-APR-1999; 99US-0129845P.  
R 19-APR-1999; 99US-0130077P.  
R 21-APR-1999; 99US-0130449P.  
R 23-APR-1999; 99US-0130510P.  
R 28-APR-1999; 99US-0130891P.  
R 30-APR-1999; 99US-0131449P.  
R 30-APR-1999; 99US-0132048P.  
R 30-APR-1999; 99US-0132407P.  
R 04-MAY-1999; 99US-0132484P.  
R 05-MAY-1999; 99US-0132485P.  
R 06-MAY-1999; 99US-0132486P.  
R 06-MAY-1999; 99US-0132487P.  
R 07-MAY-1999; 99US-0132863P.  
R 11-MAY-1999; 99US-0134256P.  
R 14-MAY-1999; 99US-0134218P.  
R 14-MAY-1999; 99US-0134219P.  
R 14-MAY-1999; 99US-0134221P.  
R 14-MAY-1999; 99US-0134370P.  
R 18-MAY-1999; 99US-0134768P.  
R 19-MAY-1999; 99US-0134941P.  
R 20-MAY-1999; 99US-0135124P.  
R 21-MAY-1999; 99US-0135353P.  
R 24-MAY-1999; 99US-0135623P.  
R 25-MAY-1999; 99US-0136021P.  
R 27-MAY-1999; 99US-0136392P.  
R 28-MAY-1999; 99US-0136782P.  
R 01-JUN-1999; 99US-0137222P.  
R 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140934P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142154P.  
PR 05-JUL-1999; 99US-0142390P.  
PR 06-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142928P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145513P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.

		Matches	4; Conservative	0; Mismatches	0; Indels	0; Gaps	0
QY	Db	2 VAEF 5 45 VAEF 48					
RESULT 64							
AAG60683							
ID AAG60683 standard; protein; 59 AA.							
XX	AC	AAG60683;					
XX	DT	18-OCT-2000 (first entry)					
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 78631.					
KW	KW	Protein identification; signal transduction pathway; metabolic pathway;					
KW	KW	hybridisation assay; genetic mapping; gene expression control; promoter;					
XX	XX	termination sequence.					
OS	OS	Arabidopsis thaliana.					
XX	PN	EP1033405-A2.					
XX	PD	06-SEP-2000.					
XX	PF	25-FEB-2000; 2000EP-00301439.					
XX	PR	25-FEB-1999; 99US-0121825P.					
PR	PR	05-MAR-1999; 99US-0123180P.					
PR	PR	09-MAR-1999; 99US-0123548P.					
PR	PR	23-MAR-1999; 99US-0125788P.					
PR	PR	25-MAR-1999; 99US-0126264P.					
PR	PR	29-MAR-1999; 99US-0126785P.					
PR	PR	01-APR-1999; 99US-0127462P.					
PR	PR	06-APR-1999; 99US-0128234P.					
PR	PR	08-APR-1999; 99US-0128714P.					
PR	PR	16-APR-1999; 99US-0129845P.					
PR	PR	19-APR-1999; 99US-0130077P.					
PR	PR	21-APR-1999; 99US-0130449P.					
PR	PR	23-APR-1999; 99US-0130510P.					
PR	PR	23-APR-1999; 99US-0130891P.					
PR	PR	28-APR-1999; 99US-0131449P.					
PR	PR	30-APR-1999; 99US-0132048P.					
PR	PR	30-APR-1999; 99US-0132407P.					
PR	PR	04-MAY-1999; 99US-0132484P.					
PR	PR	05-MAY-1999; 99US-0132485P.					
PR	PR	06-MAY-1999; 99US-0132486P.					
PR	PR	07-MAY-1999; 99US-0132487P.					
PR	PR	11-MAY-1999; 99US-0132863P.					
PR	PR	14-MAY-1999; 99US-0134218P.					
PR	PR	14-MAY-1999; 99US-0134219P.					
PR	PR	14-MAY-1999; 99US-0134221P.					
PR	PR	14-MAY-1999; 99US-0134370P.					
PR	PR	18-MAY-1999; 99US-0134768P.					
PR	PR	19-MAY-1999; 99US-0134941P.					
PR	PR	20-MAY-1999; 99US-0135124P.					
PR	PR	21-MAY-1999; 99US-0135353P.					
PR	PR	24-MAY-1999; 99US-0135629P.					
PR	PR	24-MAY-1999; 99US-0136021P.					
PR	PR	27-MAY-1999; 99US-0136392P.					
PR	PR	28-MAY-1999; 99US-0136782P.					
PR	PR	01-JUN-1999; 99US-0137222P.					
PR	PR	03-JUN-1999; 99US-0137528P.					
PR	PR	04-JUN-1999; 99US-0137502P.					
PR	PR	07-JUN-1999; 99US-0137724P.					
PR	PR	08-JUN-1999; 99US-0138094P.					
PR	PR	10-JUN-1999; 99US-0138540P.					
PR	PR	10-JUN-1999; 99US-0138847P.					
PR	PR	14-JUN-1999; 99US-0139119P.					
PR	PR	16-JUN-1999; 99US-0139452P.					
Query Match 95.0%; Score 19; DB 3; Length 59;							
Best Local Similarity 100.0%; Pred. No. 8.2e+02;							
PR	PR	06-AUG-1999; 99US-0147303P.					
PR	PR	06-AUG-1999; 99US-0147416P.					
PR	PR	09-AUG-1999; 99US-0147493P.					
PR	PR	09-AUG-1999; 99US-0147935P.					
PR	PR	10-AUG-1999; 99US-0148171P.					
PR	PR	11-AUG-1999; 99US-0148319P.					
PR	PR	12-AUG-1999; 99US-0148341P.					
PR	PR	13-AUG-1999; 99US-0148565P.					
PR	PR	13-AUG-1999; 99US-0148684P.					
PR	PR	16-AUG-1999; 99US-0149368P.					
PR	PR	17-AUG-1999; 99US-0149175P.					
PR	PR	18-AUG-1999; 99US-0149426P.					
PR	PR	20-AUG-1999; 99US-0149722P.					
PR	PR	20-AUG-1999; 99US-0149723P.					
PR	PR	22-AUG-1999; 99US-0149929P.					
PR	PR	23-AUG-1999; 99US-0149902P.					
PR	PR	23-AUG-1999; 99US-0149930P.					
PR	PR	25-AUG-1999; 99US-0150566P.					
PR	PR	26-AUG-1999; 99US-0150884P.					
PR	PR	27-AUG-1999; 99US-0151065P.					
PR	PR	27-AUG-1999; 99US-0151066P.					
PR	PR	27-AUG-1999; 99US-0151080P.					
PR	PR	30-AUG-1999; 99US-0151303P.					
PR	PR	31-AUG-1999; 99US-0151438P.					
PR	PR	01-SEP-1999; 99US-0151930P.					
PR	PR	01-SEP-1999; 99US-0152263P.					
PR	PR	10-SEP-1999; 99US-0153070P.					
PR	PR	13-SEP-1999; 99US-0153758P.					
PR	PR	15-SEP-1999; 99US-0154018P.					
PR	PR	16-SEP-1999; 99US-0154039P.					
PR	PR	20-SEP-1999; 99US-0154779P.					
PR	PR	22-SEP-1999; 99US-0155139P.					
PR	PR	23-SEP-1999; 99US-0155486P.					
PR	PR	24-SEP-1999; 99US-0155559P.					
PR	PR	28-SEP-1999; 99US-0156458P.					
PR	PR	29-SEP-1999; 99US-0156596P.					
PR	PR	04-OCT-1999; 99US-0157117P.					
PR	PR	05-OCT-1999; 99US-0157753P.					
PR	PR	06-OCT-1999; 99US-0157865P.					
PR	PR	07-OCT-1999; 99US-0158029P.					
PR	PR	08-OCT-1999; 99US-0158322P.					
PR	PR	12-OCT-1999; 99US-0158369P.					
PR	PR	13-OCT-1999; 99US-0159293P.					
PR	PR	13-OCT-1999; 99US-0159294P.					
PR	PR	13-OCT-1999; 99US-0159295P.					
PR	PR	14-OCT-1999; 99US-0159329P.					
PR	PR	14-OCT-1999; 99US-0159330P.					
PR	PR	14-OCT-1999; 99US-0159331P.					
PR	PR	14-OCT-1999; 99US-0159337P.					
PR	PR	14-OCT-1999; 99US-0159638P.					
PR	PR	18-OCT-1999; 99US-0159584P.					
PR	PR	21-OCT-1999; 99US-0160741P.					
PR	PR	21-OCT-1999; 99US-0160767P.					
PR	PR	21-OCT-1999; 99US-0160768P.					
PR	PR	21-OCT-1999; 99US-0160770P.					
PR	PR	21-OCT-1999; 99US-0160814P.					
PR	PR	21-OCT-1999; 99US-0160815P.					
PR	PR	22-OCT-1999; 99US-0160980P.					
PR	PR	22-OCT-1999; 99US-0160981P.					
PR	PR	22-OCT-1999; 99US-0160989P.					
PR	PR	25-OCT-1999; 99US-0161404P.					
PR	PR	25-OCT-1999; 99US-0161405P.					
PR	PR	25-OCT-1999; 99US-0161406P.					
PR	PR	26-OCT-1999; 99US-0161359P.					
PR	PR	26-OCT-1999; 99US-0161360P.					
PR	PR	26-OCT-1999; 99US-0161361P.					
PR	PR	26-OCT-1999; 99US-0161920P.					
PR	PR	28-OCT-1999; 99US-0161922P.					
PR	PR	28-OCT-1999; 99US-0161933P.					
PR	PR	29-OCT-1999; 99US-0162142P.					

R 16-JUN-1999; 99US-0139453P.  
R 17-JUN-1999; 99US-0139492P.  
R 18-JUN-1999; 99US-0139454P.  
R 18-JUN-1999; 99US-0139455P.  
R 18-JUN-1999; 99US-0139456P.  
R 18-JUN-1999; 99US-0139457P.  
R 18-JUN-1999; 99US-0139458P.  
R 18-JUN-1999; 99US-0139459P.  
R 18-JUN-1999; 99US-0139460P.  
R 18-JUN-1999; 99US-0139461P.  
R 18-JUN-1999; 99US-0139462P.  
R 18-JUN-1999; 99US-0139463P.  
R 18-JUN-1999; 99US-0139750P.  
R 18-JUN-1999; 99US-0139763P.  
R 21-JUN-1999; 99US-0139817P.  
R 22-JUN-1999; 99US-0139899P.  
R 23-JUN-1999; 99US-0140353P.  
R 23-JUN-1999; 99US-0140354P.  
R 24-JUN-1999; 99US-0140695P.  
R 24-JUN-1999; 99US-0140823P.  
R 28-JUN-1999; 99US-0140991P.  
R 30-JUN-1999; 99US-0141287P.  
R 01-JUL-1999; 99US-0141842P.  
R 01-JUL-1999; 99US-0142254P.  
R 02-JUL-1999; 99US-0142055P.  
R 06-JUL-1999; 99US-0142290P.  
R 08-JUL-1999; 99US-0142803P.  
R 09-JUL-1999; 99US-0142920P.  
R 12-JUL-1999; 99US-0142977P.  
R 13-JUL-1999; 99US-0143542P.  
R 14-JUL-1999; 99US-0143624P.  
R 15-JUL-1999; 99US-0144005P.  
R 16-JUL-1999; 99US-0144085P.  
R 16-JUL-1999; 99US-0144086P.  
R 19-JUL-1999; 99US-0144325P.  
R 19-JUL-1999; 99US-0144331P.  
R 19-JUL-1999; 99US-0144332P.  
R 19-JUL-1999; 99US-0144333P.  
R 19-JUL-1999; 99US-0144334P.  
R 19-JUL-1999; 99US-0144335P.  
R 20-JUL-1999; 99US-0144352P.  
R 20-JUL-1999; 99US-0144632P.  
R 20-JUL-1999; 99US-0144884P.  
R 21-JUL-1999; 99US-0144814P.  
R 21-JUL-1999; 99US-0145086P.  
R 21-JUL-1999; 99US-0145088P.  
R 22-JUL-1999; 99US-0145085P.  
R 22-JUL-1999; 99US-0145087P.  
R 22-JUL-1999; 99US-0145089P.  
R 22-JUL-1999; 99US-0145192P.  
R 23-JUL-1999; 99US-0145145P.  
R 23-JUL-1999; 99US-0145218P.  
R 23-JUL-1999; 99US-0145224P.  
R 26-JUL-1999; 99US-0145276P.  
R 27-JUL-1999; 99US-0145913P.  
R 27-JUL-1999; 99US-0145918P.  
R 27-JUL-1999; 99US-0145919P.  
R 28-JUL-1999; 99US-0145951P.  
R 02-AUG-1999; 99US-0146386P.  
R 02-AUG-1999; 99US-0146388P.  
R 02-AUG-1999; 99US-0146389P.  
R 03-AUG-1999; 99US-0147038P.  
R 04-AUG-1999; 99US-0147204P.  
R 04-AUG-1999; 99US-0147302P.  
R 05-AUG-1999; 99US-0147192P.  
R 05-AUG-1999; 99US-0147260P.  
R 06-AUG-1999; 99US-0147303P.  
R 06-AUG-1999; 99US-0147416P.  
R 09-AUG-1999; 99US-0147493P.  
R 09-AUG-1999; 99US-0147935P.  
R 10-AUG-1999; 99US-0148171P.  
R 11-AUG-1999; 99US-0148319P.  
R 12-AUG-1999; 99US-0148341P.

PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 23-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 08-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159325P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 95.0%; Score 19; DB 3; Length 59;  
Best Local Similarity 100.0%; Pred.No. 8.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
Db 45 VAEF 48

## RESULT 65

ABB42044  
ID ABB42044 standard; peptide; 60 AA.

XX AC ABB42044;

XX DT 04-FEB-2002 (first entry)

DE DE Peptide #9550 encoded by human foetal liver single exon probe.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

XX PN WO200157277-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000669.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human fetal liver.

XX PS Claim 27; SEQ ID NO 34679; 639pp + Sequence Listing; English.

XX CC The invention relates to a single exon nucleic acid probe for measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human fetal liver. The present sequence is a peptide encoded by a single exon nucleic acid probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 60 AA;

Query Match 95.0%; Score 19; DB 4; Length 60;

Best Local Similarity 100.0%; Pred. No. 8.4e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5

DB 8 VAEF 11

## RESULT 66

AAM35846  
ID AAM35846 standard; protein; 60 AA.

XX AC AAM35846;

XX DT 17-OCT-2001 (first entry)

DE DE Peptide #983 encoded by probe for measuring placental gene expression.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

XX KW genetic disorder.

XX OS Homo sapiens.

XX

PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000663.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human placenta.

XX PS Claim 27; SEQ ID NO 36115; 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENP: see AA1315-AA157546). The present sequence is a peptide encoded by one such probe. The probes are useful for producing a microarray for predicting, measuring and displaying gene expression in samples derived from human placenta. The probes are useful for antenatal diagnosis of human genetic disorders

XX SQ Sequence 60 AA;

Query Match 95.0%; Score 19; DB 4; Length 60;

Best Local Similarity 100.0%; Pred. No. 8.4e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5

DB 8 VAEF 11

## RESULT 67

AAM75737

ID AAM75737 standard; protein; 60 AA.

XX AC AAM75737;

XX DT 06-NOV-2001 (first entry)

XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 36043.

XX KW Human; bone marrow expressed exon; gene expression analysis; probe;

XX KW microarray; cancer; leukaemia; lymphoma; myeloma.

XX OS Homo sapiens.

XX PN WO200157276-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000668.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI; 2001-488900/53.

Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human bone marrow.

Example 4; SEQ ID NO 36043; 658pp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human bone marrow. They can be used to measure gene expression in bone marrow samples, which may enable the improved diagnosis and treatment of cancers such as lymphoma, leukaemia and myeloma. The present sequence is a protein encoded by one of the probes of the invention

Sequence 60 AA;

Query Match 95.0%; Score 19; DB 4; Length 60;  
Best Local Similarity 100.0%; Pred. No. 8.4e-02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
|||||  
8 VAEF 11

RESULT 68

AM62925  
D AM62925 standard; protein; 60 AA.

AM62925;

05-NOV-2001 (first entry)

Human brain expressed single exon probe encoded protein SEQ ID NO: 35030.

Human; brain expressed exon; gene expression analysis; probe; microarray; Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.

Homo sapiens.

WO200157275-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US000667.

04-FEB-2000; 2000US-0180312P.

26-MAY-2000; 2000US-0207456P.

30-JUN-2000; 2000US-00608408.

03-AUG-2000; 2000US-00632366.

21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.

04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-483446/52.

Single exon nucleic acid probes for analyzing gene expression in human brains.

Example 4; SEQ ID NO 35030; 650pp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system

diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention

Sequence 60 AA;

Query Match 95.0%; Score 19; DB 4; Length 60;  
Best Local Similarity 100.0%; Pred. No. 8.4e-02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
|||||  
8 VAEF 11

RESULT 69

ABG57475  
ID ABG57475 standard; peptide; 60 AA.

ABG57475;

25-FEB-2003 (first entry)

Human liver peptide, SEQ ID NO 36123.

Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia; coronary heart disease.

Homo sapiens.

WO200157273-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US000664.

04-FEB-2000; 2000US-0180312P.

26-MAY-2000; 2000US-0207456P.

30-JUN-2000; 2000US-00608408.

03-AUG-2000; 2000US-00632366.

21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.

04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488998/53.

Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human adult liver.

Claim 27; SEQ ID NO 36123; 658pp; English.

The invention relates to a single exon nucleic acid probe (SEN) (I) for measuring human gene expression in a sample derived from human adult liver, comprising one of 13109 defined nucleotide sequences given in the specification (or complements/ fragments). The probe hybridises at high stringency to a nucleic acid molecule expressed in the human adult liver. (I) may be used for predicting, measuring and displaying gene expression in samples derived from human adult liver. The genes identified may be involved in genetic liver diseases such as cirrhosis, hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is associated with coronary heart disease. ABG47348-ABG59930 represent human liver single exon encoded peptides of the invention. Note: The sequence information for this patent does not appear in the printed specification but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 60 AA;

Query Match 95.0%; Score 19; DB 4; Length 60;

```

Best Local Similarity 100.0%; Pred. No. 8.4e+02; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0;

DY 2 VAEF 5
   ||||
Db 8 VAEF 11

RESULT 70
AA17262
ID AAV17262 standard; peptide: 62 AA.
AC AAV17262;
XX 09-AUG-1999 (first entry)
XX HLH domain of drosophila E-spl m5.
XX ORF polypeptide; dermatomyositis; polymyositis; inclusion body myositis;
XX sarcoid myopathy; AZT myopathy; myocardial infarction; ischaemia; ERK-6;
XX extracellular signal-regulated kinase; reperfusion.
XX Drosophila melanogaster.
XX WO9927099-A1.
XX 03-JUN-1999.
XX 23-NOV-1998; 98WO-US025046.
XX 25-NOV-1997; 97US-0066848P.
XX (SUGEN-) SUGEN INC.
XX Ullrich A, Giot J;
XX WPI; 1999-357837/30.
XX Nucleic acid encoding ORF polypeptide.
XX Disclosure; Fig 2; 79pp; English.
XX The invention relates to a human ORF polypeptide, a substrate for
CC extracellular signal-regulated kinase, ERK-6. Substances that modulate
CC the activity of the ORF polypeptide can be used to treat diseases
CC selected from dermatomyositis, polymyositis, inclusion body myositis,
CC sarcoid myopathy, AZT myopathy, myocardial infarction, and ischaemia/
CC reperfusion. The probes and antibodies can be used to detect the presence
CC of ORF in a sample
XX SQ Sequence 62 AA;
Query Match 95.0%; Score 19; DB 2; Length 62;
Best Local Similarity 100.0%; Pred. No. 8.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
   ||||
Db 29 VAEF 32

RESULT 71
AAW91247
ID AAW91247 standard; protein; 62 AA.
XX AAW91247;
XX 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen SEQ ID NO:18840.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis.
XX KW

```

```

XX OS Homo sapiens.
XX PN WO200157182-A2.
XX PD 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001354.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180828P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214866P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216800P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.

```

R 26-SEP-2000; 2000US-0235484P.  
R 27-SEP-2000; 2000US-0235834P.  
R 27-SEP-2000; 2000US-0235836P.  
R 23-SEP-2000; 2000US-0236527P.  
R 29-SEP-2000; 2000US-0236367P.  
R 29-SEP-2000; 2000US-0236368P.  
R 29-SEP-2000; 2000US-0236369P.  
R 29-SEP-2000; 2000US-0236370P.  
R 02-OCT-2000; 2000US-0236802P.  
R 02-OCT-2000; 2000US-0237037P.  
R 02-OCT-2000; 2000US-0237038P.  
R 02-OCT-2000; 2000US-0237039P.  
R 02-OCT-2000; 2000US-0237040P.  
R 13-OCT-2000; 2000US-0239935P.  
R 13-OCT-2000; 2000US-0239937P.  
R 20-OCT-2000; 2000US-0240960P.  
R 20-OCT-2000; 2000US-0241221P.  
R 20-OCT-2000; 2000US-0241785P.  
R 20-OCT-2000; 2000US-0241786P.  
R 20-OCT-2000; 2000US-0241787P.  
R 20-OCT-2000; 2000US-0241808P.  
R 20-OCT-2000; 2000US-0241809P.  
R 20-OCT-2000; 2000US-0241826P.  
R 01-NOV-2000; 2000US-0244817P.  
R 08-NOV-2000; 2000US-0246474P.  
R 08-NOV-2000; 2000US-0246475P.  
R 08-NOV-2000; 2000US-0246476P.  
R 08-NOV-2000; 2000US-0246477P.  
R 08-NOV-2000; 2000US-0246478P.  
R 08-NOV-2000; 2000US-0246523P.  
R 08-NOV-2000; 2000US-0246524P.  
R 08-NOV-2000; 2000US-0246525P.  
R 08-NOV-2000; 2000US-0246526P.  
R 08-NOV-2000; 2000US-0246527P.  
R 08-NOV-2000; 2000US-0246528P.  
R 08-NOV-2000; 2000US-0246532P.  
R 08-NOV-2000; 2000US-0246609P.  
R 08-NOV-2000; 2000US-0246610P.  
R 08-NOV-2000; 2000US-0246611P.  
R 08-NOV-2000; 2000US-0246613P.  
R 17-NOV-2000; 2000US-0249207P.  
R 17-NOV-2000; 2000US-0249208P.  
R 17-NOV-2000; 2000US-0249209P.  
R 17-NOV-2000; 2000US-0249210P.  
R 17-NOV-2000; 2000US-0249211P.  
R 17-NOV-2000; 2000US-0249212P.  
R 17-NOV-2000; 2000US-0249213P.  
R 17-NOV-2000; 2000US-0249214P.  
R 17-NOV-2000; 2000US-0249215P.  
R 17-NOV-2000; 2000US-0249216P.  
R 17-NOV-2000; 2000US-0249217P.  
R 17-NOV-2000; 2000US-0249218P.  
R 17-NOV-2000; 2000US-0249244P.  
R 17-NOV-2000; 2000US-0249245P.  
R 17-NOV-2000; 2000US-0249264P.  
R 17-NOV-2000; 2000US-0249265P.  
R 17-NOV-2000; 2000US-0249297P.  
R 17-NOV-2000; 2000US-0249399P.  
R 17-NOV-2000; 2000US-0249300P.  
R 01-DEC-2000; 2000US-0250160P.  
R 01-DEC-2000; 2000US-0250391P.  
R 05-DEC-2000; 2000US-0251030P.  
R 05-DEC-2000; 2000US-0251988P.  
R 05-DEC-2000; 2000US-0256719P.  
R 06-DEC-2000; 2000US-0251479P.  
R 08-DEC-2000; 2000US-0251856P.  
R 08-DEC-2000; 2000US-0251868P.  
R 08-DEC-2000; 2000US-0251869P.  
R 08-DEC-2000; 2000US-0251989P.  
R 08-DEC-2000; 2000US-0251990P.  
R 11-DEC-2000; 2000US-0254097P.  
R 05-JAN-2001; 2001US-0259678P.  
XX

PA (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
DR N-PSDB; AAK64028.  
XX Nucleic acids encoding human immune/haematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX Claim 11; SEQ ID NO 18840; 3071pp + Sequence Listing; English.  
PS AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to prevent, the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX SQ Sequence 62 AA;  
Query Match 95.0%; Score 19; DB 4; Length 62;  
Best Local Similarity 100.0%; Pred. No. 8.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 VAEF 5  
Db |||||  
25 VAEF 28  
RESULT 72  
ABM65667  
ID ABM65667 standard; protein; 62 AA.  
XX ABM65667;  
AC  
XX 20-OCT-2003 (first entry)  
DT  
XX Propionibacterium acnes immunogenic polypeptide #30343.  
DE  
XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
KW immunostimulant; immune response; vaccine; immunogenic.  
KW  
XX Propionibacterium acnes.  
OS  
XX WO2003033515-A1.  
XX 24-APR-2003.  
PD  
XX 11-OCT-2002; 2002WO-US032727.  
PF  
XX 15-OCT-2001; 2001US-00978825.  
PR  
XX (CORI-) CORIXA CORP.  
PA Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
XX Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
PI Barth B, Vallieve-Douglas J;  
XX WPI; 2003-381789/36.  
DR New Propionibacterium acnes polypeptides and polynucleotides encoding the  
XX

PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.  
 XX  
 PS Claim 7; SEQ ID NO 30343; 1481pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a specifically claimed P. acnes polypeptide which is  
 CC thought to contain an immunogenic region. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX  
 SQ Sequence 62 AA;

Query Match 95.0%; Score 19; DB 6; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 8.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 |||||  
 Db 8 VAEF 11

RESULT 73  
 AAC09517  
 ID AAC09517 standard; protein; 64 AA.  
 AC AAC09517;  
 XX  
 XX 06-NOV-2001 (first entry)  
 XX  
 XX Human polypeptide SEQ ID NO 23409.  
 XX  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200164835-A2.  
 PN  
 XX  
 XX 07-SEP-2001.  
 PD  
 XX 26-FEB-2001; 2001WO-US004927.  
 PF  
 XX 28-FEB-2000; 2000US-00515126.  
 PR  
 XX 18-MAY-2000; 2000US-00577409.  
 PR  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.  
 DR N-PSDB; AA183446.  
 XX  
 XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
 PT and treating e.g. leukemia, inflammation and immune disorders.  
 PT  
 XX Claim 20; SEQ ID NO 23409; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AA179941-AA193841) and  
 CC the encoded proteins (AAO00010-AAO3910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 64 AA;

Query Match 95.0%; Score 19; DB 4; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 |||||  
 Db 54 VAEF 57

RESULT 74  
 AAY48352  
 ID AAY48352 standard; protein; 67 AA.  
 XX  
 AC AAY48352;  
 XX  
 XX 08-DEC-1999 (first entry)  
 DT  
 XX Human prostate cancer-associated protein 49.  
 DE  
 XX Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;  
 KW cancer; tissue specificity; human.  
 XX  
 XX Homo sapiens.  
 OS  
 XX DE19811194-A1.  
 PN  
 XX 16-SEP-1999.  
 PD  
 XX 10-MAR-1998; 98DE-01011194.  
 PF  
 XX 10-MAR-1998; 98DE-01011194.  
 PR  
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 XX  
 XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;  
 PI  
 XX WPI; 1999-519629/44.  
 DR N-PSDB; AA233459.  
 DR  
 XX New nucleic acid expressed at high level in normal prostatic tissue and  
 PT encoded polypeptides, used to treat cancer and screen for therapeutic  
 PT agents.  
 PT  
 XX Claim 22; 143; 194pp; German.  
 PS  
 XX This invention describes novel nucleic acid sequences (A) that are  
 CC expressed at high level in normal prostatic tissue. Polypeptides (I)  
 CC encoded by (A) are used: (a) for identifying agents for treatment of

prostatic cancer and (b) for therapy of prostate cancer, optionally where expressed by gene therapy methods. (A) is also used to isolate full-length genes (for gene therapy) and for recombinant production of (I), which can be used to raise specific antibodies. (A) are identified by assembly of ESTs (expressed sequence tags) before these are analyzed for expression pattern (tissue specificity). This approach eliminates many of the false results, as regards tissue specificity, associated with known methods that use single (usually short) ESTs. AA48304-Y48456 represent peptides encoded by the expressed sequence tags described in the method of the invention

XX Sequence 67 AA;  
 Query Match 95.0%; Score 19; DB 2; Length 67;  
 Best Local Similarity 100.0%; Pred. No. 9.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

YY 2 VAEF 5  
 |||||  
 18 VAEF 21

RESULT 75  
 ABP31047  
 ID ABP31047 standard; protein; 68 AA.  
 AC ABP31047;  
 CX  
 JT 08-JUL-2002 (first entry)  
 CX  
 DE Human ORF20 protein, SEQ ID NO:40.  
 CX  
 W Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
 W disease monitoring; cytokine; cell proliferation; cell differentiation;  
 W immune modulation; haematopoiesis regulation; tissue growth;  
 W angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
 W thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 W behaviour; cancer; proliferative disorder; neurological disorder;  
 W cardiovascular disease; immune system disorder; organ transplantation;  
 W tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 W hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
 W vasotrophic; antipsoriatic; antidiabetic; cytostatic; nootropic;  
 W neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
 W cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
 W dermatological; analgesic; viricide; antibacterial; fungicide.  
 CX  
 XS Homo sapiens.  
 CX  
 WN WO200190366-A2.  
 CX  
 XD 29-NOV-2001.  
 CX  
 XF 24-MAY-2001; 2001WO-US017076.  
 CX  
 XR 24-MAY-2000; 2000US-0206690P.  
 CX  
 XA (CURA-) CURAGEN CORP.  
 CX  
 XT Leach MD, Shimkets RA;  
 CX  
 XR WPI; 2002-106200/14.  
 CX  
 XR N-PSDB; ABN75073.  
 CX  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation.  
 CX  
 PS Claim 10; Page 280; 2508pp; English.  
 CX  
 XC Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 XC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 XC ABN75957 represent cDNAs encoding them. The invention also encompasses

polypeptides at least 80% identical to the ORF1-ORF4534 (collectively referred to as ORFX) proteins, polynucleotides at least 85% identical to the ORFX nucleic acid sequences, vectors and host cells comprising ORFX polynucleotides, the recombinant production of ORFX proteins, antibodies specific for ORFX proteins, methods of detecting ORFX polynucleotides and polypeptides, methods of screening for modulators of ORFX expression or activity, and methods of screening individuals for a predisposition to an ORFX-associated disorder. The ORFX proteins of the invention have a wide range of biological activities, such as cytokine, cell proliferation, cell differentiation, immune modulation, haematopoiesis regulation, tissue growth, angiogenesis, activin or inhibin activity, chemotactic/chemokinetic activity, haemostatic activity, thrombolytic activity, receptor/ligand, antiinflammatory activity, tumour inhibition activity, CC of bodily characteristics, fertility and behaviour. ORFX proteins, CC nucleic acids and antibodies may be used in the treatment of cancers, CC other proliferative disorders such as psoriasis and benign tumours, CC neurological disorders such as epilepsy and Alzheimer's disease, CC cardiovascular diseases, immune system disorders, disorders related to organ transplantation, disorders of tissue growth and regeneration, CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester storage disease, and infectious diseases caused by viral, bacterial, CC fungal and other pathogens. ORFX nucleic acids may also be used as a source of primers and probes, in the detection of ORFX genomic sequences or transcripts, in the identification and cloning of homologous CC sequences, in genetic diagnosis, and in forensic biology. The ORFX CC nucleic acids may additionally be used to produce transgenic animals CC which may be useful for studying the function and/or activity of ORFX CC protein, and in drug screening. The ORFX proteins may also be used as CC immunogens to generate specific antibodies, which are useful in the CC diagnosis, treatment and monitoring of ORFX-associated diseases

XX  
 SQ Sequence 68 AA;

Query Match 95.0%; Score 19; DB 5; Length 68;

Best Local Similarity 100.0%; Pred. No. 9.6e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5

DB 3 VAEF 6

Search completed: May 24, 2004, 17:38:08

Job time : 53.7857 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: May 24, 2004, 17:32:13 ; Search time 12.5 Seconds  
(without alignments)  
39.477 Million cell updates/sec

Title: US-09-594-978A-3

Perfect score: 20

Sequence: 1 XVAEF 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : PIR\_78:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	95.0	20	S39049	cytotoxin-binding
2	19	95.0	33	G54625	hypothetical prote
3	19	95.0	48	E64618	hypothetical prote
4	19	95.0	52	A82375	hypothetical prote
5	19	95.0	54	E85745	unknown protein en
6	19	95.0	59	A69386	conserved hypothet
7	19	95.0	59	E71349	probable preprotel
8	19	95.0	63	UC4002	carbomycin resista
9	19	95.0	63	G64007	hypothetical prote
10	19	95.0	64	I51350	transferrin - Atla
11	19	95.0	74	D37045	hypothetical prote
12	19	95.0	75	S07277	gene D protein - p
13	19	95.0	79	T18034	hypothetical prote
14	19	95.0	79	T27605	hypothetical prote
15	19	95.0	84	F83922	hypothetical prote
16	19	95.0	87	A38725	transferrin - shee
17	19	95.0	89	H64116	ribosomal protein
18	19	95.0	89	F66818	hypothetical prote
19	19	95.0	89	T15018	hypothetical prote
20	19	95.0	96	AB3053	hypothetical prote
21	19	95.0	97	AC1397	conserved hypothet
22	19	95.0	99	B83244	conserved hypothet
23	19	95.0	101	H90297	conserved hypothet
24	19	95.0	106	T10097	nifx protein - Met
25	19	95.0	108	C64490	hypothetical prote
26	19	95.0	109	E70949	probable regulator
27	19	95.0	111	T17582	hypothetical prote
28	19	95.0	112	B83431	type III export pr
29	19	95.0	114	A55872	kedarcidin [valida

30	19	95.0	114	2	D97414	hypothetical prote
31	19	95.0	115	2	D75218	hypothetical prote
32	19	95.0	116	1	IPAF	insulin precursor
33	19	95.0	118	2	T48682	hypothetical prote
34	19	95.0	119	2	F95327	hypothetical prote
35	19	95.0	122	2	F82861	conjugal transfer
36	19	95.0	125	2	B69129	gamma-carboxymuon
37	19	95.0	125	2	D70750	hypothetical prote
38	19	95.0	126	2	B81368	hypothetical prote
39	19	95.0	127	2	D70759	hypothetical prote
40	19	95.0	127	2	AB1732	hypothetical prote
41	19	95.0	128	2	A54797	ileal lipid-bindin
42	19	95.0	129	2	T02012	probable cathepsin
43	19	95.0	130	2	D72348	conserved hypothet
44	19	95.0	131	2	G75208	hypothetical prote
45	19	95.0	132	2	JQ2270	hypothetical 15.2K
46	19	95.0	132	2	AG1996	hypothetical prote
47	19	95.0	133	2	AH0137	conserved hypothet
48	19	95.0	133	2	H75623	conserved hypothet
49	19	95.0	133	2	A84221	riboflavin synthas
50	19	95.0	134	1	WMEC15	ybGC protein - Esc
51	19	95.0	134	2	C90725	hypothetical prote
52	19	95.0	134	2	D85576	hypothetical prote
53	19	95.0	134	2	AD0592	conserved hypothet
54	19	95.0	136	2	S76416	hypothetical prote
55	19	95.0	137	2	B81355	hypothetical prote
56	19	95.0	138	2	F82106	flagellar basal-bo
57	19	95.0	139	2	A69052	riboflavin synthas
58	19	95.0	139	2	A83975	hypothetical prote
59	19	95.0	140	2	B69323	hypothetical prote
60	19	95.0	142	2	JQ2279	low-temperature re
61	19	95.0	142	2	JQ2280	riboflavin synthas
62	19	95.0	143	2	H69515	probable transposo
63	19	95.0	143	2	T36978	ribosomal protein
64	19	95.0	145	1	B41715	hypothetical prote
65	19	95.0	145	2	S77943	conserved hypothet
66	19	95.0	146	2	F89837	hypothetical prote
67	19	95.0	147	2	C84294	hypothetical prote
68	19	95.0	147	2	T34803	conserved hypothet
69	19	95.0	148	2	B81788	hypothetical prote
70	19	95.0	148	2	S10655	hypothetical prote
71	19	95.0	149	2	AC1137	ribose 5-phosphate
72	19	95.0	150	2	AB2857	pseudoazurin limpo
73	19	95.0	150	2	AB1089	protein gp35 from
74	19	95.0	150	2	AB1452	protein gp35 from
75	19	95.0	151	2	H87263	hypothetical prote
76	19	95.0	152	2	S04623	superoxide dismuta
77	19	95.0	154	2	B90184	hypothetical prote
78	19	95.0	155	2	I64130	PAI cross-reacting
79	19	95.0	156	2	G81283	ankyrin-repeat con
80	19	95.0	159	2	T17848	hypothetical prote
81	19	95.0	161	2	AB0981	probable acetyltra
82	19	95.0	163	2	A24504	salivary glue prot
83	19	95.0	163	2	G70524	hypothetical prote
84	19	95.0	164	2	AD1229	B. subtilis Y61B p
85	19	95.0	164	2	A69293	hypothetical prote
86	19	95.0	166	2	S50509	hypothetical prote
87	19	95.0	168	2	A48866	hypothetical 19.5K
88	19	95.0	168	2	AB0192	probable C4-dicarb
89	19	95.0	169	2	T36387	probable acetyltra
90	19	95.0	169	2	AB2739	acetyltransferase
91	19	95.0	170	2	AC2917	invasion protein A
92	19	95.0	172	2	B75041	hypothetical prote
93	19	95.0	172	2	I49449	hemopoietic-specif
94	19	95.0	173	2	T22350	hypothetical prote
95	19	95.0	174	2	B70879	hypothetical prote
96	19	95.0	175	1	G71120	hypothetical prote
97	19	95.0	175	2	AC2648	conserved hypothet
98	19	95.0	175	2	I39055	Bcl-2 related - hu
99	19	95.0	176	2	C27805	ferritin chain M -
100	19	95.0	178	1	B69389	hypothetical prote
101	19	95.0	178	2	A85933	transcription fact
102	19	95.0	178	2	S03629	neurogenic gene co

103	19	95.0	179	2	S34345	hypothetical prote	176	19	95.0	219	2	H69199	conserved hypothet
104	19	95.0	180	2	D87931	probable acetyltra	177	19	95.0	219	2	AE1506	weakly carboxylest
105	19	95.0	180	2	C85358	hypothetical prote	178	19	95.0	220	2	T10279	protein tyrosine p
106	19	95.0	180	2	AI0676	probable exported	179	19	95.0	220	2	B84281	riboflavin-specifi
107	19	95.0	181	2	AG2202	cobinamide kinase	180	19	95.0	221	2	C83947	hypothetical prote
108	19	95.0	183	2	S13186	plasma retinol-bin	181	19	95.0	223	2	D69311	conserved hypothet
109	19	95.0	184	2	A90903	probable ante-term	182	19	95.0	223	2	T35665	hypothetical prote
110	19	95.0	184	2	F90873	probable anti-termi	183	19	95.0	223	2	AC0890	probable lipoprote
111	19	95.0	184	2	D71333	hypothetical prote	184	19	95.0	224	2	S69643	hypothetical prote
112	19	95.0	185	2	S43729	H+-transporting tw	185	19	95.0	224	2	AH1187	hypothetical prote
113	19	95.0	185	2	S39317	replication initia	186	19	95.0	226	2	AG3528	hypothetical prote
114	19	95.0	185	2	S39318	replication initia	187	19	95.0	226	2	A75290	chloramphenicol ac
115	19	95.0	185	2	D75635	resolvase - deinoc	188	19	95.0	228	2	S48116	integral membrane
116	19	95.0	185	2	B97430	hypothetical prote	189	19	95.0	228	2	AD0693	pyridoxal kinase (
117	19	95.0	187	2	S16314	photosynthetic rea	190	19	95.0	228	2	D30857	hypothetical prote
118	19	95.0	187	2	AG2721	conserved hypothet	191	19	95.0	228	2	T10399	hypothetical prote
119	19	95.0	187	2	A95896	conserved hypothet	192	19	95.0	228	2	AH0081	probable exported
120	19	95.0	188	2	AB0132	conserved hypothet	193	19	95.0	229	2	F96721	probable glutathio
121	19	95.0	188	2	A71286	conserved hypothet	194	19	95.0	229	2	E70340	glutaredoxin-like
122	19	95.0	188	2	H85713	unknown protein en	195	19	95.0	230	2	E97440	DNA-3-methyladenin
123	19	95.0	189	2	T29159	hypothetical prote	196	19	95.0	230	2	S73229	ribosomal protein
124	19	95.0	189	2	H97633	pseudotaurin precu	197	19	95.0	231	2	AF1144	phosphoglycerate m
125	19	95.0	190	2	A29413	ubiquinol-cytochro	198	19	95.0	231	2	AF1503	phosphoglycerate m
126	19	95.0	191	2	A82392	transcription regu	199	19	95.0	231	2	C83179	conserved hypothet
127	19	95.0	192	2	D97503	hypothetical prote	200	19	95.0	231	2	AF3380	hypothetical prote
128	19	95.0	193	1	C64328	conserved hypothet	201	19	95.0	232	2	B70653	hypothetical prote
129	19	95.0	194	1	RPECR5	resolvase - Escher	202	19	95.0	232	2	B87520	hypothetical prote
130	19	95.0	194	2	C84033	hypothetical prote	203	19	95.0	234	2	F96620	hypothetical prote
131	19	95.0	194	2	F97691	hypothetical prote	204	19	95.0	234	2	AC2185	hypothetical prote
132	19	95.0	194	2	G87453	RNA polymerase sig	205	19	95.0	234	2	E91119	hypothetical prote
133	19	95.0	195	2	B86204	hypothetical prote	206	19	95.0	234	2	D85964	hypothetical prote
134	19	95.0	196	1	BVECAU	transcription regu	207	19	95.0	234	2	S22360	hypothetical prote
135	19	95.0	196	2	A41853	hexose phosphate t	208	19	95.0	235	2	B97753	(p)ppGpp 3-pyropho
136	19	95.0	196	2	AH0963	two-component syst	209	19	95.0	235	2	C95848	probable transcrip
137	19	95.0	196	2	F91204	transcription regu	210	19	95.0	236	2	A98351	agropine synthesis
138	19	95.0	196	2	H86050	hypothetical prote	211	19	95.0	236	2	AE2931	agropine synthesis
139	19	95.0	196	2	A64888	probable resolvase	212	19	95.0	236	2	B70728	hypothetical prote
140	19	95.0	196	2	D64909	probable resolvase	213	19	95.0	236	2	A84686	hypothetical prote
141	19	95.0	196	2	T21347	hypothetical prote	214	19	95.0	236	2	T50908	hypothetical prote
142	19	95.0	197	2	C86748	hypothetical prote	215	19	95.0	237	2	F83699	two-component resp
143	19	95.0	198	2	S48749	cytochrome P460 pr	216	19	95.0	237	2	C97085	hypothetical prote
144	19	95.0	199	1	VAHU	plasma retinol-bin	217	19	95.0	238	2	C86656	glycerol uptake fa
145	19	95.0	199	2	H90206	amidotransferase h	218	19	95.0	239	2	D81218	conserved hypothet
146	19	95.0	200	2	AB0620	probable bacteriop	219	19	95.0	239	2	F81795	probable periplasm
147	19	95.0	201	1	VARB	retinol-binding pr	220	19	95.0	239	2	H51955	hypothetical prote
148	19	95.0	201	2	A39486	plasma retinol-bin	221	19	95.0	239	2	S31033	gene 88 protein -
149	19	95.0	201	2	I46257	retinol binding pr	222	19	95.0	240	2	D69261	sulfate ABC transp
150	19	95.0	201	2	T00799	hypothetical prote	223	19	95.0	240	2	B84606	probable ATP synth
151	19	95.0	202	2	T35126	hypothetical prote	224	19	95.0	241	2	T14808	hypothetical prote
152	19	95.0	202	2	S77556	hypothetical prote	225	19	95.0	242	2	B70366	hypothetical prote
153	19	95.0	202	2	AH1169	hypothetical prote	226	19	95.0	242	2	G97959	hypothetical prote
154	19	95.0	204	2	C97866	hypothetical prote	227	19	95.0	247	2	S43728	H+-transporting tw
155	19	95.0	204	2	G89753	protein filic7.1 li	228	19	95.0	248	2	AF0177	probable beta-keto
156	19	95.0	205	2	S30739	hypothetical prote	229	19	95.0	248	2	T12632	probable channel prot
157	19	95.0	205	2	H91225	hypothetical prote	230	19	95.0	248	2	T01648	probable tonoplast
158	19	95.0	205	2	F86072	hypothetical prote	231	19	95.0	248	2	F72604	probable high-affi
159	19	95.0	205	2	G03942	mitotic feedback c	232	19	95.0	248	2	T18315	hypothetical prote
160	19	95.0	208	2	A48642	hypothetical prote	233	19	95.0	249	2	G81693	serine/threonine p
161	19	95.0	209	2	AH2658	3-methyladenine-DN	234	19	95.0	249	2	B69343	conserved hypothet
162	19	95.0	209	2	H97519	hypothetical prote	235	19	95.0	249	2	T59088	hypothetical prote
163	19	95.0	212	2	A83521	conserved hypothet	236	19	95.0	250	2	JQ1012	TobR87-18C protein
164	19	95.0	213	2	G82067	hypothetical prote	237	19	95.0	250	2	S13719	probable membrane
165	19	95.0	213	2	AF2551	hypothetical prote	238	19	95.0	250	2	S51781	integral membrane
166	19	95.0	213	2	A84248	transcription regu	239	19	95.0	250	2	D64580	hypothetical prote
167	19	95.0	214	2	T30310	probable GTP bindi	240	19	95.0	251	2	F69494	(R)-hydroxyglutary
168	19	95.0	215	2	A13072	transcription regu	241	19	95.0	252	2	C50422	hypothetical prote
169	19	95.0	215	2	F84889	hypothetical prote	242	19	95.0	252	2	F87575	ABC transporter, p
170	19	95.0	218	2	H83485	conserved hypothet	243	19	95.0	252	2	G44020	hypothetical prote
171	19	95.0	218	2	T03287	osmotin protein ho	244	19	95.0	253	1	S17711	probable dehydroge
172	19	95.0	218	2	D71392	coat protein - cuc	245	19	95.0	253	2	C65067	2-deoxy-D-gluconat
173	19	95.0	218	2	AB0254	probable funarylac	246	19	95.0	253	2	C31091	2-deoxy-D-gluconat
174	19	95.0	218	2	T47889	hypothetical prote	247	19	95.0	253	2	F85936	2-deoxy-D-gluconat
175	19	95.0	219	2	T35128	hypothetical prote	248	19	95.0	253	2	AD0210	2-deoxy-D-gluconat

249	19	95.0	253	2	B72552	probable ATP-depen
250	19	95.0	254	2	F90087	ATP-dependent Clp
251	19	95.0	254	2	T02078	photosystem II oxy
252	19	95.0	254	2	AE0161	deoxyribose operon
253	19	95.0	255	2	S12985	phosphoprotein pho
254	19	95.0	256	2	T52130	probable copper/zinc
255	19	95.0	257	2	T35273	probable enoyl coA
256	19	95.0	257	2	S56108	26S proteasome reg
257	19	95.0	258	2	B83275	conserved hypochet
258	19	95.0	258	2	F69479	nicotinate-nucleot
259	19	95.0	258	2	T36359	hypothetical prote
260	19	95.0	259	2	S75539	hypothetical prote
261	19	95.0	259	2	G81427	periplasmic protei
262	19	95.0	260	2	AF2054	hypothetical prote
263	19	95.0	261	2	AD2895	short-chain dehydr
264	19	95.0	261	2	G97670	probable oxidoredu
265	19	95.0	261	2	T00100	hypothetical prote
266	19	95.0	261	2	G84010	hypothetical prote
267	19	95.0	261	2	H98213	transcription repr
268	19	95.0	262	2	B72333	bacteriocin - Ther
269	19	95.0	267	2	AC0132	probable aldo/keto
270	19	95.0	267	2	A70490	matrilysin (EC 3.4
271	19	95.0	267	2	T24397	hypothetical prote
272	19	95.0	267	2	A87404	ribosomal protein
273	19	95.0	267	2	A13121	ABC transporter, m
274	19	95.0	267	2	F98165	hypothetical prote
275	19	95.0	269	1	R5BYL3	ribosomal protein
276	19	95.0	269	2	B44999	major intrinsic pr
277	19	95.0	269	2	A41616	erythrocyte integr
278	19	95.0	269	2	UC1320	water channel prot
279	19	95.0	269	2	I52366	uterine water chan
280	19	95.0	269	2	C71228	hypothetical prote
281	19	95.0	269	2	A56487	signal recognition
282	19	95.0	271	2	UC2348	water channel prot
283	19	95.0	272	2	B99937	conserved hypochet
284	19	95.0	274	2	S37705	amylC protein - The
285	19	95.0	274	2	H88690	protein F4H10.8 [
286	19	95.0	274	2	C83709	hypothetical prote
287	19	95.0	274	2	H93737	acetyl-CoA carboxy
288	19	95.0	275	2	B97260	dihydroneopterin a
289	19	95.0	276	2	AG1842	uroporphyrinogen-I
290	19	95.0	276	2	T19845	hypothetical prote
291	19	95.0	277	2	JW0102	azaarene carbazole
292	19	95.0	277	2	A72420	conserved hypochet
293	19	95.0	278	2	E81983	hypothetical prote
294	19	95.0	278	2	A81040	conserved hypochet
295	19	95.0	278	2	A81040	probable nicotinam
296	19	95.0	279	2	E71293	prephenate dehydra
297	19	95.0	279	2	S52582	prephenate dehydra
298	19	95.0	279	2	F86842	hypothetical prote
299	19	95.0	279	2	T22051	hypothetical prote
300	19	95.0	279	2	D84185	site-specific DNA-
301	19	95.0	280	2	S71506	transcription regu
302	19	95.0	280	2	AC3579	hypothetical prote
303	19	95.0	281	2	C86221	hypothetical prote
304	19	95.0	281	2	E84637	fof1-type ATP synt
305	19	95.0	282	2	F97252	tRNA-pseudouridine
306	19	95.0	282	2	B70200	hypothetical prote
307	19	95.0	282	2	H83675	2-hydroxyruconate-
308	19	95.0	284	2	T31275	probable aldose re
309	19	95.0	284	2	T37996	hypothetical prote
310	19	95.0	285	2	B70703	envelope-like prot
311	19	95.0	286	2	A84494	2-hydroxyruconate-
312	19	95.0	286	2	JC5419	pyridoxal kinase (
313	19	95.0	286	2	A10288	glutamine-tRNA lig
314	19	95.0	287	2	S07563	probable pyridoxal
315	19	95.0	287	2	F64920	pyridoxal kinase 2
316	19	95.0	287	2	A90922	pyridoxal kinase 2
317	19	95.0	287	2	B85770	hypothetical prote
318	19	95.0	288	2	T35229	glycine hydroxymet
319	19	95.0	288	2	A13400	hypothetical prote
320	19	95.0	289	1	G64167	ABC transporter, m
321	19	95.0	290	2	AD2870	probable aliphatic
322	19	95.0	290	2	F97645	

395	19	95.0	320	2	T28992	hypotheical prote	468	19	95.0	352	2	AC1560	two-component sens
396	19	95.0	322	2	D89785	hypotheical prote	469	19	95.0	352	2	AE1202	hypulin-like pept
397	19	95.0	323	2	T09550	phosphoprotein pho	470	19	95.0	353	2	D97395	hypulin-like pept
398	19	95.0	323	2	G84325	hypotheical prote	471	19	95.0	355	2	AG1877	hypulin-like pept
399	19	95.0	324	2	E84305	hypotheical prote	472	19	95.0	355	2	AI3624	spemidine/putresc
400	19	95.0	325	2	GC7553	brain mitochondria	473	19	95.0	357	2	AC1276	aminopeptidase hom
401	19	95.0	326	2	G69366	homoserine dehydro	474	19	95.0	357	2	AC1639	aminopeptidase hom
402	19	95.0	326	2	T05094	peroxidase homolog	475	19	95.0	357	2	AC1639	hypotheical prote
403	19	95.0	326	2	T19070	hypotheical prote	476	19	95.0	359	2	T02011	probable cathepsin
404	19	95.0	327	2	AC3482	sulfate-binding pr	477	19	95.0	359	2	D95316	probable ABC-trans
405	19	95.0	327	2	E95968	hypotheical expor	478	19	95.0	359	2	E95382	probable ABC trans
406	19	95.0	328	2	E75564	ABC transporter, A	479	19	95.0	360	2	D69025	pleiotropic regula
407	19	95.0	328	2	A89867	hypotheical prote	480	19	95.0	360	2	AB3078	hypotheical prote
408	19	95.0	328	2	C70368	conserved hypotet	481	19	95.0	360	2	G98208	probable ATP-bindi
409	19	95.0	328	2	H84548	hypotheical prote	482	19	95.0	361	2	D98243	hypotheical prote
410	19	95.0	329	2	AD2629	33 kda chaperonin,	483	19	95.0	361	2	D98243	opuaa (Ar234619) (
411	19	95.0	329	2	T50117	mo25 homolog (impo	484	19	95.0	361	2	AE1979	ABC transporter AT
412	19	95.0	329	2	T23218	hypotheical prote	485	19	95.0	361	2	E96743	probable deoxyguan
413	19	95.0	330	1	H64404	hypotheical prote	486	19	95.0	362	2	C97448	hypotheical prote
414	19	95.0	330	2	A83417	probable oxidoredu	487	19	95.0	363	2	E95371	probable oxidoredu
415	19	95.0	331	2	T35499	probable D-lactate	488	19	95.0	363	2	AE1134	acyl-CoA dehydrog
416	19	95.0	331	2	AE23928	dehydrogenase Atu3	489	19	95.0	363	2	C81888	probable acyl-CoA
417	19	95.0	331	2	D82972	probable lipolytic	490	19	95.0	364	2	G70364	conserved hypotet
418	19	95.0	331	2	AE2666	divalent cation tr	491	19	95.0	365	2	S76544	hypotheical prote
419	19	95.0	332	2	D75067	probable dehydrog	492	19	95.0	366	1	G49964	photosynthetic rea
420	19	95.0	333	2	C71165	probable dehydrog	493	19	95.0	366	2	TS0891	cytochrome subunit
421	19	95.0	333	2	A11881	collagenase Prtc (	494	19	95.0	366	2	G84219	NADH-dependent fla
422	19	95.0	334	1	SL1318	transcription regu	495	19	95.0	368	2	G70462	succinate-CoA liga
423	19	95.0	334	2	T49195	hypotheical prote	496	19	95.0	368	2	G97291	hypotheical prote
424	19	95.0	334	2	A95951	hypotheical prote	497	19	95.0	368	2	S76552	ABC-type transport
425	19	95.0	335	2	S42807	H8R203J protein -	498	19	95.0	368	2	C83898	response regulator
426	19	95.0	336	2	E95906	probable dehydrog	499	19	95.0	368	2	A97189	lps biosynthesis p
427	19	95.0	336	2	AE0337	probable aspartate	500	19	95.0	369	2	S56638	mitogen-activated
428	19	95.0	336	2	AH3569	ABC transporter AT	501	19	95.0	369	2	A55124	RNA polymerase sig
429	19	95.0	336	2	C87623	cobalamin biosynth	502	19	95.0	369	2	C97934	rna polymerase sig
430	19	95.0	336	2	T00571	delichyl-phosphate	503	19	95.0	369	2	B83571	probable ATP-bandi
431	19	95.0	337	2	T35862	probable secreted	504	19	95.0	369	2	S77028	protein kinase, 41
432	19	95.0	337	2	T22757	hypotheical prote	505	19	95.0	369	2	TS1477	glutamine-rich pro
433	19	95.0	338	2	AC2938	2-hydroxyacid-fam	506	19	95.0	370	2	F36819	C14 protein - rabb
434	19	95.0	338	2	H98343	hypotheical prote	507	19	95.0	370	2	F95927	probable mandelate
435	19	95.0	338	2	E75301	conserved hypotet	508	19	95.0	370	2	JC7927	platelet-derived g
436	19	95.0	339	2	SL7930	transcription init	509	19	95.0	371	2	T09622	protein kinase MMK
437	19	95.0	339	2	JC7509	glycoprotein Vi-1	510	19	95.0	371	2	F86625	Grp-binding protei
438	19	95.0	339	2	T25630	hypotheical prote	511	19	95.0	371	2	D95000	Grp-binding protei
439	19	95.0	340	2	JC1125	phosphotransferase	512	19	95.0	371	2	D97872	conserved hypotet
440	19	95.0	340	2	T32646	hypotheical prote	513	19	95.0	371	2	AH0686	hypotheical prote
441	19	95.0	340	2	E83146	alpha 4 protein -	514	19	95.0	372	2	AD0075	probable transport
442	19	95.0	340	2	S71223	xyloglucan endo-1,	515	19	95.0	372	2	AC0253	ribonuclease iii (
443	19	95.0	341	2	AH0087	sigma-54 transcrip	516	19	95.0	375	2	H98185	diptidase (impor
444	19	95.0	342	2	S29894	strictosidine synt	517	19	95.0	375	2	T16116	membrane dipepida
445	19	95.0	342	2	AD3258	cobs protein (impo	518	19	95.0	376	2	A48352	genome polyprotein
446	19	95.0	342	2	E86446	probable endoxylg	519	19	95.0	377	2	D95861	probable ABC trans
447	19	95.0	343	2	C89886	protein C50H11.7 (	520	19	95.0	377	2	D95861	tela protein - pla
448	19	95.0	343	2	D84390	sulfate transport	521	19	95.0	378	1	B38178	mannosyl transfera
449	19	95.0	343	2	RWRTC2	T-cell surface gly	522	19	95.0	379	2	E92409	guanine tRNA-ribos
450	19	95.0	344	1	A98354	hypotheical prote	523	19	95.0	379	2	H82284	NAD(PAD)-depend
451	19	95.0	344	2	B28967	T-cell surface gly	524	19	95.0	379	2	G97201	NADH-rubredoxin ox
452	19	95.0	344	2	I49585	CD2 antigen protei	525	19	95.0	379	2	AD3390	phnM protein (impo
453	19	95.0	344	2	AI1902	4-hydroxyphenylpr	526	19	95.0	379	2	AE3225	ABC transporter, m
454	19	95.0	344	2	S01325	strictosidine synt	527	19	95.0	379	2	B45333	cystathionine gamm
455	19	95.0	344	2	F90239	anthranilate phosp	528	19	95.0	380	2	D71973	probable cystathio
456	19	95.0	345	2	E83927	hypotheical prote	529	19	95.0	380	2	B45333	442K curved dna-bi
457	19	95.0	345	2	A45885	MHC class I histoc	530	19	95.0	381	2	AG3304	pleiotropic regula
458	19	95.0	346	2	C95404	hypotheical prote	531	19	95.0	381	2	D72777	probable S2P metal
459	19	95.0	346	2	C90590	hypotheical prote	532	19	95.0	383	2	D86821	hydroxymethylgluta
460	19	95.0	346	2	D63573	immunogenic protei	533	19	95.0	384	2	T16447	hypotheical prote
461	19	95.0	347	2	D63573	flagellar motor sw	534	19	95.0	385	2	H84411	hypotheical prote
462	19	95.0	347	2	AF2645	flagellar motor sw	535	19	95.0	385	2	D54074	phosphoglycerate k
463	19	95.0	347	2	F97427	sulfate transport	536	19	95.0	386	2	B97411	31K chaperonin (he
464	19	95.0	348	2	E69162	probable O-sialogl	537	19	95.0	386	2	AC3190	hypotheical prote
465	19	95.0	349	2	H72714	hypotheical prote	538	19	95.0	387	2	E70894	probable me8a prot
466	19	95.0	351	2	D84541	strictosidine synt	539	19	95.0	388	2	F87208	cystathionine (gam
467	19	95.0	352	2	S22464		540	19	95.0	388	2	F87208	

541	19	95.0	388	2	T23670	hypothetical prote	614	19	95.0	427	2	G98259	methionine gamma-l
542	19	95.0	389	2	T15102	hypothetical prote	615	19	95.0	428	2	F69050	glycine hydroxymet
543	19	95.0	389	2	G42455	NADH dehydrogenase	616	19	95.0	429	2	G72323	phosphopyruvate hy
544	19	95.0	390	2	E92722	conserved hypochet	617	19	95.0	429	2	S30049	transcription fact
545	19	95.0	390	2	H75023	nicotinate phospho	618	19	95.0	429	2	AC2227	hypothetical prote
546	19	95.0	390	2	A84226	farnesyl-diphospha	619	19	95.0	430	2	AC2227	O-acetylhomoserine
547	19	95.0	390	2	T25996	hypothetical prote	620	19	95.0	430	2	B95892	probable ABC trans
548	19	95.0	391	2	D6291	hypothetical prote	621	19	95.0	431	2	C69087	dihydroliipoamide d
549	19	95.0	391	2	C70972	probable trehalose	622	19	95.0	431	2	C69087	enolase [imported]
550	19	95.0	392	2	B4017	probable RNA methy	623	19	95.0	431	2	T21594	hypothetical prote
551	19	95.0	392	2	S22580	telomere-binding p	624	19	95.0	431	2	D70185	3-hydroxy-3-methyl
552	19	95.0	393	2	E95261	serine proteinase	625	19	95.0	431	2	S37775	filamin, muscle -
553	19	95.0	393	2	C89961	hypothetical prote	626	19	95.0	432	1	S15203	glycine hydroxymet
554	19	95.0	394	2	B70411	fimbrial assembly	627	19	95.0	432	2	S08277	cyclin A - human
555	19	95.0	394	2	S13133	probable efflux pu	628	19	95.0	432	2	AB2403	cytochrome P450 Rv
556	19	95.0	394	2	G76331	hypothetical prote	629	19	95.0	433	1	B70677	probable GTP bindi
557	19	95.0	395	2	S38812	cyclin A - chicken	630	19	95.0	433	2	A70465	glycine hydroxymet
558	19	95.0	395	2	T20724	hypothetical prote	631	19	95.0	434	2	S30334	probable flagellum
559	19	95.0	397	2	H84225	acyl-CoA dehydroge	632	19	95.0	434	2	D71480	3-hydroxy-3-methyl
560	19	95.0	397	2	H84225	serine proteinase	633	19	95.0	435	2	G69466	transcription fact
561	19	95.0	397	2	A72315	hypothetical prote	634	19	95.0	437	1	A31752	tetracyclorolipol
562	19	95.0	398	2	C95201	hydroxymethylgluta	635	19	95.0	437	2	C64113	high-affinity gluc
563	19	95.0	398	2	A98068	hypothetical prote	636	19	95.0	437	2	JC4988	transcription fact
564	19	95.0	399	2	A84972	nicotinate phospho	637	19	95.0	437	2	S42111	probable aminotran
565	19	95.0	399	2	H87342	glycosyl transfera	638	19	95.0	438	2	E95383	probable aminotran
566	19	95.0	399	2	F83796	multidrug-efflux t	639	19	95.0	438	2	A98161	hypothetical prote
567	19	95.0	399	2	AC2221	hypothetical prote	640	19	95.0	438	2	H86006	hypothetical prote
568	19	95.0	400	2	T03460	probable leucine/i	641	19	95.0	439	2	AB0997	high-affinity gluc
569	19	95.0	402	1	A70707	cytochrome P450 Rv	642	19	95.0	439	2	AC0462	hypothetical prote
570	19	95.0	402	1	H82813	ABC transporter, A	643	19	95.0	439	2	AC2948	glycerol-3-phospha
571	19	95.0	403	2	H82440	NADH dehydrogenase	644	19	95.0	440	2	G98334	glycerol-3-phospha
572	19	95.0	403	2	G84222	probable metallope	645	19	95.0	440	2	AC2985	conserved hypochet
573	19	95.0	405	2	A83084	cyclin A - bovine	646	19	95.0	440	2	D98398	probable sugar iso
574	19	95.0	406	2	S24788	serine proteinase	647	19	95.0	441	2	C87233	probable cell inva
575	19	95.0	406	2	F94152	tryptophan oxygene	648	19	95.0	442	2	T44655	O-acetylhomoserine
576	19	95.0	406	2	G20222	hypothetical prote	649	19	95.0	443	2	A99657	hypothetical prote
577	19	95.0	406	2	T30748	hypothetical prote	650	19	95.0	443	2	A85508	hypothetical prote
578	19	95.0	407	2	T71679	protein kinase, 48	651	19	95.0	443	2	T32088	hypothetical prote
579	19	95.0	408	2	T08069	hypothetical prote	652	19	95.0	444	2	C70444	biotin carboxylase
580	19	95.0	408	2	B97741	hypothetical prote	653	19	95.0	445	2	T16025	hypothetical prote
581	19	95.0	409	2	A04992	multidrug transloc	654	19	95.0	445	2	S20336	finger protein XFG
582	19	95.0	409	2	T02776	Y4dM protein - Rhi	655	19	95.0	445	2	AB0970	undulin 2 - human
583	19	95.0	409	2	F86151	P22M8.10 protein -	656	19	95.0	446	2	C82140	C4-dicarboxylate t
584	19	95.0	410	2	B48585	transcription fact	657	19	95.0	447	2	C96497	glyceraldehyde-3-p
585	19	95.0	410	2	A48585	transcription fact	658	19	95.0	448	2	AF0437	conserved hypochet
586	19	95.0	411	2	H83355	hypothetical prote	659	19	95.0	449	2	D70726	probable gabr - My
587	19	95.0	411	2	AB0199	peptidase T (EC 3.	660	19	95.0	449	2	C70846	probable metC prot
588	19	95.0	412	1	RNECTA	tRNA adenylyltrans	661	19	95.0	449	2	Ar2646	chemotaxis motD pr
589	19	95.0	412	1	S51760	ferredoxin reducta	662	19	95.0	449	2	B97428	chemotaxis motD pr
590	19	95.0	412	2	C91121	tRNA nucleotidyl t	663	19	95.0	450	2	T39433	glucoamylase precu
591	19	95.0	412	2	B85965	tRNA nucleotidyl t	664	19	95.0	450	2	F82402	permease VCA0904 l
592	19	95.0	412	2	T21419	hypothetical prote	665	19	95.0	450	2	T21593	hypothetical prote
593	19	95.0	413	2	C95160	phosphoenolpyruvat	666	19	95.0	450	2	C86466	hypothetical prote
594	19	95.0	413	2	T48364	transcription fact	667	19	95.0	453	2	S37935	conserved hypochet
595	19	95.0	415	2	A91920	glutamy1-tRNA redu	668	19	95.0	453	2	C83008	probable purine-bi
596	19	95.0	415	2	C91183	glutamy1-tRNA redu	669	19	95.0	455	2	B71335	hypothetical prote
597	19	95.0	415	2	I51637	cyclin A2 - Africa	670	19	95.0	456	2	T46986	DNA-directed DNA p
598	19	95.0	416	2	D56231	diaminopimelate de	671	19	95.0	456	2	AF0240	repsA protein - St
599	19	95.0	416	2	F98095	diaminopimelate de	672	19	95.0	458	2	S75328	hypothetical prote
600	19	95.0	418	2	H72203	hypothetical prote	673	19	95.0	458	2	AT2218	hypothetical prote
601	19	95.0	418	2	T35993	probable aminotran	674	19	95.0	460	2	AT2218	argininosuccinate
602	19	95.0	419	2	B85035	hypothetical prote	675	19	95.0	461	1	WZTRRS	argininosuccinate
603	19	95.0	420	2	S65084	finger protein XFG	676	19	95.0	461	2	B71672	malonyl-CoA decarb
604	19	95.0	421	2	A90003	hypothetical prote	677	19	95.0	462	2	A33313	gene XGF 5.1c prot
605	19	95.0	421	2	T47393	hypothetical prote	678	19	95.0	462	2	D97826	malonate hydratase
606	19	95.0	422	2	C83063	glutamy1-tRNA redu	679	19	95.0	463	1	WZTRRS	argininosuccinate
607	19	95.0	422	2	S38501	cyclin A2 - mouse	680	19	95.0	464	1	WZTRRS	argininosuccinate
608	19	95.0	422	2	S37280	hypothetical prote	681	19	95.0	465	2	P70837	probable sulfatase
609	19	95.0	423	2	T19145	citrate (si)-synth	682	19	95.0	466	2	T36212	replication initia
610	19	95.0	424	2	I139506	hypothetical prote	683	19	95.0	467	2	AT2231	nicotinamide nucle
611	19	95.0	424	2	H87520	hypothetical prote	684	19	95.0	467	2	G82697	hypothetical prote
612	19	95.0	426	2	H75407	transcription term	685	19	95.0	468	2	S39832	probable phosphor
613	19	95.0	427	2	AB3025	methionine gamma-l	686	19	95.0	469	1	KCPGI	interstitial colla

687	19	95.0	469	2	S74825	probable Rieske ir	760	19	95.0	513	2	A85697	Na+/H+ antiporter,
688	19	95.0	470	2	S50083	phorolase - short	761	13	95.0	514	2	G64864	Na+/H+-exchanging
689	19	95.0	471	2	S27676	mannose-1-phosphat	762	13	95.0	514	2	S46733	hypothetical prote
690	19	95.0	471	2	C82917	preprotein translo	763	19	95.0	514	2	T10559	hypothetical prote
691	19	95.0	471	2	T04738	hypothetical prote	764	19	95.0	514	2	AG0723	regulator of intra
692	19	95.0	472	2	B47677	phosphomannomutase	765	19	95.0	517	2	T29852	hypothetical prote
693	19	95.0	472	2	S22621	conserved hypotet	766	19	95.0	517	2	T27101	hypothetical prote
694	19	95.0	473	2	D69503	hypothetical prote	767	19	95.0	518	1	A27705	alpha-amylase (EC
695	19	95.0	473	2	T24197	hydrogenase large	768	19	95.0	521	2	T46687	p-cresol methylhyd
696	19	95.0	474	2	S75530	proline dipeptidas	769	19	95.0	522	2	A31556	glucose transpor
697	19	95.0	474	2	AG3210	long-chain-fatty-a	770	19	95.0	522	2	B64067	Na+/H+-exchanging
698	19	95.0	475	2	B49681	hypothetical prote	771	19	95.0	523	2	S06920	glucose transport
699	19	95.0	475	2	G87292	starch synthase (E	772	19	95.0	524	1	T40127	IMP dehydrogenase
700	19	95.0	476	2	C64119	hypothetical prote	773	19	95.0	524	2	A31318	glucose transport
701	19	95.0	476	2	C96577	hypothetical prote	774	19	95.0	524	2	T06134	hypothetical prote
702	19	95.0	476	2	S71360	Noea protein - phi	775	19	95.0	524	2	T08931	hypothetical prote
703	19	95.0	476	2	H95313	adenyl cyclase-a	776	19	95.0	526	2	D95205	hypothetical prote
704	19	95.0	477	2	JC4386	adenyl cyclase-a	777	19	95.0	528	2	T31459	protochlorophyllid
705	19	95.0	477	2	J38409	hypothetical prote	778	19	95.0	528	2	JC4814	Na+/H+-exchanging
706	19	95.0	478	2	T23792	hypothetical prote	779	19	95.0	529	2	AF3059	cytochrome d oxida
707	19	95.0	478	2	S15312	rfbm protein - Sal	780	19	95.0	529	2	B98227	cytochrome d oxida
708	19	95.0	479	2	AH0765	mannose-1-phosphat	781	19	95.0	532	2	T07903	tubulin delta chai
709	19	95.0	481	2	B86367	protein F6F24.16	782	19	95.0	533	2	S52046	deoxyribodipyrimid
710	19	95.0	481	2	H96729	probable alanine a	783	19	95.0	536	2	C82433	methyl-accepting c
711	19	95.0	481	2	A70190	hypothetical prote	784	19	95.0	537	2	T09710	4-coumarate-CoA li
712	19	95.0	483	2	T10095	nife protein - Met	785	19	95.0	537	2	AF0575	2,3-dihydroxybenzo
713	19	95.0	483	2	T47422	cellulase-like pro	786	19	95.0	537	2	H87502	conserved hypotet
714	19	95.0	483	2	D82446	probable formate t	787	19	95.0	538	2	AF3281	IMP cyclohydrolase
715	19	95.0	484	2	T36302	probable 2-dehydro	788	19	95.0	538	2	AF2923	bifunctional purin
716	19	95.0	484	2	E86416	unknown protein, 3	789	19	95.0	540	2	S76584	hypothetical prote
717	19	95.0	485	2	F95238	PTS system, membra	790	19	95.0	543	2	H82282	probable chemotaxi
718	19	95.0	486	2	H83131	5-carboxy-2-hydrox	791	19	95.0	543	2	F83280	probable chemotaxi
719	19	95.0	486	2	C96018	probable starch sy	792	19	95.0	548	2	AB0691	fumate hydratase
720	19	95.0	486	2	T49017	hypothetical prote	793	19	95.0	548	2	AB1854	glucose-6-phosphat
721	19	95.0	486	2	T39456	zinc finger protei	794	19	95.0	548	2	C81089	glucose-6-phosphat
722	19	95.0	488	2	A95926	probable argininos	795	19	95.0	548	2	AB1854	glucose-6-phosphat
723	19	95.0	488	2	T09734	1-aminocyclopropan	796	19	95.0	548	2	T23270	acetyl choline rec
724	19	95.0	488	2	T23055	hypothetical prote	797	19	95.0	549	2	A90253	hypothetical prote
725	19	95.0	488	2	H90423	sugar transport pr	798	19	95.0	549	2	T24508	hypothetical prote
726	19	95.0	490	2	T44576	L-2,4-diaminobuty	799	19	95.0	549	2	JC4083	chaperonin - Caeno
727	19	95.0	490	2	E71486	probable 8/t prote	800	19	95.0	550	2	B82330	glucose-6-phosphat
728	19	95.0	491	2	S73053	glucose-6-phosphat	801	19	95.0	550	2	T24019	hypothetical prote
729	19	95.0	491	2	T11894	NADH2 dehydrogenas	802	19	95.0	550	2	T22557	hypothetical prote
730	19	95.0	492	2	B64642	cell division prot	803	19	95.0	551	2	E84106	regulatory protein
731	19	95.0	492	2	T01235	hypothetical prote	804	19	95.0	553	2	S03828	cytochrome d ubiqu
732	19	95.0	493	2	C71873	septum formation p	805	19	95.0	553	2	AG3604	malate synthase (E
733	19	95.0	493	2	H70573	hypothetical prote	806	19	95.0	553	1	S1FQMA	steroid monooxygen
734	19	95.0	494	2	S10134	58K mobilization p	807	19	95.0	556	2	C87472	hypothetical prote
735	19	95.0	494	2	A33952	plasmid recombinat	808	19	95.0	556	2	T16790	serine C-palmitoyl
736	19	95.0	495	2	T27693	hypothetical prote	809	19	95.0	558	2	T29503	hypothetical prote
737	19	95.0	495	2	AD2275	hypothetical prote	810	19	95.0	558	2	A75216	hypothetical prote
738	19	95.0	497	2	S27785	acetyl-CoA acetyl	811	19	95.0	559	2	T28881	hypothetical prote
739	19	95.0	497	2	AG3134	N-ethylammine ch	812	19	95.0	559	2	AB0984	probable membrane
740	19	95.0	498	2	T14236	NADH2 dehydrogenas	813	19	95.0	559	2	H82532	conserved hypotet
741	19	95.0	498	2	F88359	protein Y51HA.5 [	814	19	95.0	561	2	G81842	30S ribosomal prot
742	19	95.0	498	2	F88153	N-ethylammine ch	815	19	95.0	561	2	G81842	30S ribosomal prot
743	19	95.0	499	1	B64069	fructose phosphot	816	19	95.0	561	2	H69086	ABC transporter -
744	19	95.0	499	2	H86448	hypothetical prote	817	19	95.0	561	2	H69086	hypothetical prote
745	19	95.0	500	2	C19211	probable tail shea	818	19	95.0	562	2	F97697	methyl-accepting c
746	19	95.0	502	2	B81863	exopolysphatase	819	19	95.0	564	2	B72660	probable type II D
747	19	95.0	502	2	F81080	exopolysphatase	820	19	95.0	565	2	B72660	probable type II D
748	19	95.0	503	2	G81917	probable ubiquinon	821	19	95.0	567	2	S69779	hypothetical prote
749	19	95.0	503	2	B81194	ubiquinone biosynt	822	19	95.0	567	2	T16105	conserved hypotet
750	19	95.0	503	2	A82153	Sun/nucleolar prot	823	19	95.0	567	2	D72359	beta-glucuronidase
751	19	95.0	504	2	S27113	malonyl-CoA decarb	824	19	95.0	570	2	C90485	hypothetical prote
752	19	95.0	504	2	A84244	probable signaling	825	19	95.0	571	2	T06737	hypothetical prote
753	19	95.0	504	2	T16903	hypothetical prote	826	19	95.0	576	1	S35853	proline-tRNA ligas
754	19	95.0	506	2	C70468	phosphoribosylamin	827	19	95.0	576	2	B71959	acetylactate synch
755	19	95.0	506	2	T31831	hypothetical prote	828	19	95.0	577	2	C69059	proline-tRNA ligas
756	19	95.0	507	2	AD0475	probable magnesium	829	19	95.0	577	2	F64549	probable peptidase
757	19	95.0	508	2	G98102	hypothetical prote	830	19	95.0	580	2	T30583	probable peptidase
758	19	95.0	512	2	T14638	Cytochrome P450 Cy	831	19	95.0	580	2	C86530	S1 ribosomal prote
759	19	95.0	513	2	A96839	Na+/H+ antiporter	832	19	95.0	580	2	D72093	ribosomal protein

833	19	95.0	532	2	T16104	hypothetical prote	906	19	95.0	639	2	A56126	peroxisomal target
834	19	95.0	533	2	G86481	Mutator-like trans	907	19	95.0	640	1	A55073	transforming prote
835	19	95.0	534	2	T46177	villin 3 homolog T	908	19	95.0	640	2	F75553	hypothetical prote
836	19	95.0	537	2	E82431	methy1-accepting c	909	19	95.0	641	2	AD3502	dnak protein limpo
837	19	95.0	538	2	T25248	hypothetical prote	910	19	95.0	642	2	S55521	beta-fructofuranos
838	19	95.0	538	2	A43740	DG42 protein - Afr	911	19	95.0	644	2	JQ0160	3-isopropylmalate
839	19	95.0	538	2	AB3460	hypothetical cytos	912	19	95.0	644	2	A45635	dnak-type molecula
840	19	95.0	539	2	A46459	macrophage-activat	913	19	95.0	645	2	JC5642	dnak-type molecula
841	19	95.0	539	2	S74668	hypothetical prote	914	19	95.0	645	2	T25824	hypothetical prote
842	19	95.0	539	2	B81104	nitrate/nitrite se	915	19	95.0	646	2	B86295	hypothetical prote
843	19	95.0	539	2	C81911	nitrate/nitrite se	916	19	95.0	647	2	AF1488	transcription anti
844	19	95.0	539	2	S43506	hypothetical prote	917	19	95.0	648	2	AG0314	probable thiamin p
845	19	95.0	539	2	S70524	guanine nucleotide	918	19	95.0	649	2	JN0809	drebrin E (clone g
846	19	95.0	539	2	S64923	probable membrane	919	19	95.0	650	2	F72540	hypothetical prote
847	19	95.0	538	2	T45827	pectinesterase-lik	920	19	95.0	651	2	T12083	beta-fructofuranos
848	19	95.0	538	2	S71095	hypothetical prote	921	19	95.0	651	2	JC4610	dnak-type molecula
849	19	95.0	538	2	T32430	hypothetical prote	922	19	95.0	652	2	T14763	hypothetical prote
850	19	95.0	538	2	D69331	probable DNA topoi	923	19	95.0	652	2	I39811	hypothetical prote
851	19	95.0	538	2	T01360	hypothetical prote	924	19	95.0	654	2	AD3183	parasporeal crystal
852	19	95.0	538	2	T39674	hypothetical dnaw	925	19	95.0	656	1	B70107	transferrin-bindin
853	19	95.0	538	2	F79336	protein M01G12.12	926	19	95.0	660	2	S70904	peptidase limpo
854	19	95.0	538	2	T23659	hypothetical prote	927	19	95.0	669	2	C82384	hypothetical prote
855	19	95.0	538	2	I39837	dnak-type molecula	928	19	95.0	669	2	C96630	hypothetical prote
856	19	95.0	538	2	D84687	hypothetical prote	929	19	95.0	670	2	D86457	hypothetical prote
857	19	95.0	538	2	S46833	hypothetical prote	930	19	95.0	670	2	AH425	hypothetical prote
858	19	95.0	538	2	T35710	transferase - Stre	931	19	95.0	672	2	S61157	hypothetical prote
859	19	95.0	538	2	B86744	dnak protein limpo	932	19	95.0	673	2	C83080	probable chemotaxi
860	19	95.0	538	2	S39342	dnak-type molecula	933	19	95.0	674	2	E84300	bacterio-opsin act
861	19	95.0	538	2	B95060	hypothetical prote	934	19	95.0	674	2	T22733	hypothetical prote
862	19	95.0	538	2	G97328	hypothetical prote	935	19	95.0	678	1	RDRT04	NADPH-ferrhemopro
863	19	95.0	538	2	A95898	probable sensor ki	936	19	95.0	678	2	B48089	growth arrest-spec
864	19	95.0	538	2	S41315	hypothetical prote	937	19	95.0	680	2	A42297	peptidyl-dipeptida
865	19	95.0	538	2	C89939	dnak protein limpo	938	19	95.0	680	2	AH0678	dipeptidyl carboxy
866	19	95.0	538	2	G69130	conserved hypotet	939	19	95.0	684	2	T25603	hypothetical prote
867	19	95.0	538	2	A70756	probable ABC-type	940	19	95.0	685	2	S64985	hypothetical prote
868	19	95.0	538	2	T43738	dnak-type molecula	941	19	95.0	686	2	C64428	hypothetical prote
869	19	95.0	538	2	AE1621	class I heat-shock	942	19	95.0	686	2	E86409	hypothetical prote
870	19	95.0	538	2	AE1258	class I heat-shock	943	19	95.0	688	2	D96930	hypothetical prote
871	19	95.0	538	2	A11258	protein kinase pkn	944	19	95.0	688	2	T11749	transferrin - Atla
872	19	95.0	538	2	S77221	hypothetical prote	945	19	95.0	691	2	T32748	hypothetical prote
873	19	95.0	538	2	A84411	hypothetical prote	946	19	95.0	691	2	AD2766	conserved hypotet
874	19	95.0	538	2	T06717	phosphinothricin-t	947	19	95.0	691	2	H97546	hypothetical prote
875	19	95.0	538	2	P00059	two-component sens	948	19	95.0	694	1	TFRBP	transferrin precu
876	19	95.0	538	2	B83160	parasporeal crystal	949	19	95.0	696	1	S01384	hypothetical prote
877	19	95.0	538	2	SL7402	hypothetical prote	950	19	95.0	696	2	E96831	hypothetical prote
878	19	95.0	538	2	C83389	transferrin-bindin	951	19	95.0	696	2	H83024	transferrin precu
879	19	95.0	538	2	D64107	asparagine synthas	952	19	95.0	698	1	TrHUP	asma protein VC103
880	19	95.0	538	2	D86903	probable periplasm	953	19	95.0	704	2	G82249	carbonic anhydrase
881	19	95.0	538	2	G96976	carnitine O-acetyl	954	19	95.0	704	2	I47228	phenylalanine ammo
882	19	95.0	538	2	S53369	hypothetical prote	955	19	95.0	705	2	JC5873	hypothetical prote
883	19	95.0	538	2	D71974	hypothetical prote	956	19	95.0	705	2	D86227	transferrin precu
884	19	95.0	538	2	H64390	carbon-monoxide de	957	19	95.0	706	2	S33761	phenylalanine ammo
885	19	95.0	538	2	B64534	hypothetical prote	958	19	95.0	707	2	S29029	drebrin A - rat
886	19	95.0	538	2	NPV217	nucleoside-triphos	959	19	95.0	707	2	S60588	large n antigen -
887	19	95.0	538	2	H36847	hypothetical prote	960	19	95.0	708	1	TVFPT4	phenylalanine ammo
888	19	95.0	538	2	T28539	hypothetical prote	961	19	95.0	708	2	T14295	phenylalanine ammo
889	19	95.0	538	2	C72163	OLI protein - vari	962	19	95.0	708	2	T10909	conserved hypotet
890	19	95.0	538	2	T37384	nucleoside triphos	963	19	95.0	708	2	A82195	phenylalanine ammo
891	19	95.0	538	2	T37098	transferrin-bindin	964	19	95.0	710	2	JC5872	ATP-dependent DNA
892	19	95.0	538	2	T30702	hypothetical prote	965	19	95.0	712	2	E83226	phenylalanine ammo
893	19	95.0	538	2	S32349	probable SNF2-type	966	19	95.0	712	2	T03663	phenylalanine ammo
894	19	95.0	538	2	B64112	dnak-type molecula	967	19	95.0	712	2	T01858	hemolysin secretio
895	19	95.0	538	2	B82273	dnak protein VC085	968	19	95.0	714	2	A41454	hypothetical prote
896	19	95.0	538	2	AE1881	1-deoxy-xylulose 5	969	19	95.0	714	2	T35770	fatty-acid beta-ox
897	19	95.0	538	2	TVM5MB	transforming prote	970	19	95.0	715	1	JX0199	glycyl-tRNA synthe
898	19	95.0	538	2	A7042	dnak-type molecula	971	19	95.0	715	2	G83269	conserved hypotet
899	19	95.0	538	2	B30352	dnak protein FA476	972	19	95.0	715	4	TVMSWY	transforming prote
900	19	95.0	538	2	E75044	DNA repair helicase	973	19	95.0	716	2	AH2319	glycyl-tRNA synthe
901	19	95.0	538	2	IOECDK	dnak-type molecula	974	19	95.0	716	2	D69855	conserved hypotet
902	19	95.0	538	2	F90630	heat shock protein	975	19	95.0	716	2	G82627	hypothetical prote
903	19	95.0	538	2	F85481	dnak-type molecula	976	19	95.0	721	2	A70764	probable gigX prot
904	19	95.0	538	2	A50503	dnak protein (heat	977	19	95.0	722	2	S76367	hypothetical prote
905	19	95.0	538	2	S72163	methy1-accepting c	978	19	95.0	722	2		

979 19 95.0 724 2 B85045 probable calcium c  
 980 19 95.0 725 1 Z2BP44 rapid lysais protei  
 981 19 95.0 725 2 S52990 phenylalanine ammo  
 982 19 95.0 725 2 G84787 phenylalanine ammo  
 983 19 95.0 726 2 A10147 probable ATP-depen  
 984 19 95.0 729 2 AF0808 probable membrane  
 985 19 95.0 729 2 AF0808 dipeptidyl peptida  
 986 19 95.0 738 2 A87516 hypothetical prote  
 987 19 95.0 742 2 D85882 NADH2 dehydrogenas  
 988 19 95.0 744 2 T13048 genome polypeptid  
 989 19 95.0 746 2 S31816 hypothetical prote  
 990 19 95.0 748 2 G96498 hypothetical prote  
 991 19 95.0 754 2 T25551 subtilisin-like pr  
 992 19 95.0 756 2 T04187 transforming prote  
 993 19 95.0 761 1 TVHUMB polysulfide reduct  
 994 19 95.0 762 2 S23457 aconitate hydratase  
 995 19 95.0 764 2 S49849 sensory box/G3DEF  
 996 19 95.0 768 2 H82367 yfzA protein - Esc  
 997 19 95.0 772 2 T05838 subtilisin-like pr  
 998 19 95.0 772 2 T05838 probable aminopept  
 999 19 95.0 774 2 G71308 DNA polymerase B1  
 1000 19 95.0 781 2 A69312

## ALIGNMENTS

RESULT 1  
 S39049  
 cytoxin-binding protein - rabbit  
 C/Species: Oryctolagus cuniculus (domestic rabbit)  
 C/Date: 18-Feb-1994 #sequence\_revision 19-Apr-1996 #text\_change 16-Feb-1997  
 R/Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenna  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.  
 Eur. J. Biochem. 217, 1123-1128, 1993  
 A/Title: Pseudomonas aeruginosa cytoxin-binding protein in rabbit erythrocyte membrane  
 A/Reference number: S39049, MUID:94039134; PMID:7693466  
 A/Accession: S39049  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 1-20 <LUT>  
 C/Superfamily: lens fiber membrane major intrinsic protein

Query Match 95.0%; Score 19; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 ||||  
 DB 14 VAEF 17

RESULT 2  
 G64625  
 hypothetical protein HP0847 - Helicobacter pylori (strain 26695)  
 C/Species: Helicobacter pylori  
 C/Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
 R/Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenna  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.  
 Nature 388, 539-547, 1997  
 A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A/Reference number: A64520; MUID:97394467; PMID:9252185  
 A/Accession: G64625  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-33 <TOM>  
 A/Cross-references: GB:AE000596; GB:AE000511; NID:G2313982; PIDN:AAD07920.1; PID:G231400

Query Match 95.0%; Score 19; DB 2; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 88;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 ||||  
 DB 15 VAEF 18

RESULT 3  
 E64618  
 hypothetical protein HP0789 - Helicobacter pylori (strain 26695)

C/Species: Helicobacter pylori  
 C/Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
 R/Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenna  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.  
 Nature 388, 539-547, 1997  
 A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A/Reference number: A64520; MUID:97394467; PMID:9252185  
 A/Accession: E64618  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-48 <TOM>  
 A/Cross-references: GB:AE000591; GB:AE000511; NID:G2313918; PIDN:AAD07847.1; PID:G23139

Query Match 95.0%; Score 19; DB 2; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 ||||  
 DB 15 VAEF 18

RESULT 4  
 A42375  
 hypothetical protein (lon 5' region) - Bacillus brevis

C/Species: Bacillus brevis  
 C/Date: 10-Jul-1992 #sequence\_revision 10-Jul-1992 #text\_change 15-Oct-1999  
 C/Accession: A42375; I39873  
 R/Ito, K.; Ueda, S.; Yamagata, H.  
 J. Bacteriol. 174, 2281-2287, 1992  
 A/Title: Cloning, characterization, and inactivation of the Bacillus brevis lon gene.  
 A/Reference number: A42375; MUID:92202157; PMID:1551846  
 A/Accession: A42375  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-52 <ITO>  
 A/Cross-references: GB:D00863; NID:G216293; PIDN:BAA00736.1; PID:G303507

Query Match 95.0%; Score 19; DB 2; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 ||||  
 DB 47 VAEF 50

RESULT 5

E85745  
 unknown protein encoded within prophage CP-933R [imported] - Escherichia coli (strain O  
 C/Species: Escherichia coli  
 C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C/Accession: E85745  
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimantal, E.; Potamousis, K.; Apodaca  
 Nature 409, 529-533, 2001  
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A/Reference number: A85480; MUID:21074935; PMID:11206551  
 A/Accession: E85745  
 A/Status: preliminary

```

A/Molecule type: DNA
A/Residues: 1-54 <STO>
A/Cross-references: GB:AE005174; NID:G12515374; PIDN:AA056425.1; GSPDB:GN00145; UWGP:Z23
A/Experimental source: strain O157:H7, substrain EDL933
C/Genetics:
A/Gene: Z2382

Query Match 95.0%; Score 19; DB 2; Length 54;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
DB 32 VAEF 35

RESULT 6
A69386
conserved hypothetical protein AF1090 - Archaeoglobus fulgidus
C/Species: Archaeoglobus fulgidus
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jan-2000
C/Accession: A69386
A/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
R.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
ature 390, 364-370, 1997
A/Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A/Reference number: A69250; MUID:98049343; PMID:9389475
A/Accession: A69386
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-59 <KLE>
A/Cross-references: GB:AE001028; GB:AE000782; NID:G2689351; PIDN:AA090157.1; PID:G264950
C/Superfamily: Methanococcus jannaschii hypothetical protein MJ0975

Query Match 95.0%; Score 19; DB 2; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
DB 51 VAEF 54

RESULT 7
E71349
probable preprotein translocase subunit (secE) - syphilis spirochete
C/Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C/Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C/Accession: E71349
R/Praser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwir
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A/Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A/Reference number: A71250; MUID:98332770; PMID:9665876
A/Accession: E71349
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-59 <COL>
A/Cross-references: GB:AE001205; GB:AE000520; NID:G3322501; PIDN:AA065223.1; PID:G332250
A/Experimental source: strain Nichols
C/Genetics:
A/Gene: TP0235

Query Match 95.0%; Score 19; DB 2; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
DB 32 VAEF 35

RESULT 8
JC4002
carboxymycin resistance protein carA - Streptomyces sp. (fragment)
C/Species: Streptomyces sp.
C/Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 20-Sep-1999
C/Accession: JC4002
R/Arisawa, A.; Teunekawa, H.; Okamura, K.; Okamoto, R.
Biosci. Biotechnol. Biochem. 59, 582-588, 1995
A/Title: Nucleotide sequence analysis of the carboxymycin biosynthetic genes including the
A/Reference number: JC4001; MUID:95290751; PMID:7772821
A/Accession: JC4002
A/Molecule type: DNA
A/Residues: 1-63 <ARI>
A/Cross-references: DDBJ:D30759; NID:G551628; PIDN:BA06419.1; PID:dl006989; PID:G551629
A/Note: the source was designated as Streptomyces thermotolerans
C/Genetics:
A/Gene: carA
C/Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C/Keywords: ATP
P:1-55/Domain: ATP-binding cassette homology (fragment) <ABC2>

Query Match 95.0%; Score 19; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
DB 57 VAEF 60

RESULT 9
G64007
hypothetical protein HI0451 - Haemophilus influenzae (strain Rd KW20)
C/Species: Haemophilus influenzae
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C/Accession: G64007
R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MUID:95350630; PMID:7542800
A/Accession: G64007
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-63 <TIGR>
A/Cross-references: GB:U32728; GB:L42023; NID:G1573425; PIDN:AA022109.1; PID:G1573437; T

Query Match 95.0%; Score 19; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
DB 27 VAEF 30

RESULT 10
IS1350
transferrin - Atlantic salmon (fragment)
C/Species: Salmo salar (Atlantic salmon)
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C/Accession: IS1350
R/Kvingedal, A.M.
Gene 150, 335-339, 1994
A/Title: Characterization of the 5' region of the Atlantic salmon (Salmo salar) transfer
A/Reference number: IS1350; MUID:95121925; PMID:7821802
A/Accession: IS1350
A/Status: preliminary; translated from GB/EMBL/DBJ

```

```

A;Molecule type: DNA
A;Residues: 1-64 <KVI>
A;Cross-references: GB:L26909; NID:G598395; PIDN:AA42221.1; PID:G598396
C;Genetics:
A;Gene: Tf
A;Introns: 14/1
C;Superfamily: transferrin; transferrin repeat homology

Query Match          95.0%; Score 19; DB 2; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 VAEF 5
DB      45 VAEF 48

RESULT 11
D97045
hypothetical protein CAC1179 [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: D97045
R;Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koorin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: D97045
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-74 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK79151.1; PID:G15024100; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1179

Query Match          95.0%; Score 19; DB 2; Length 74;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 VAEF 5
DB      5 VAEF 8

RESULT 12
S07277
gene D protein - phase 186 (fragment)
C;Species: phase 186
C;Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 17-Mar-1999
C;Accession: S07277; S25273
R;Kallionis, B.; Dodd, I.B.; Egan, J.B.
J. Mol. Biol. 191, 199-209, 1986
A;Title: Control of gene expression in the P2-related template coliphages. III. DNA sequ
A;Reference number: S07277; MUID:87112711; PMID:3806670
A;Accession: S07277
A;Molecule type: DNA
A;Residues: 1-75 <KAL>
A;Cross-references: EMBL:X04449; NID:g15549; PID:g15550
R;Dibbels, J.A.; Gregory, S.L.; Egan, J.B.
Mol. Microbiol. 6, 2643-2650, 1992
A;Title: Control of gene expression in the temperate coliphage 186. X. The cl repressor
A;Reference number: S25273; MUID:93078618; PMID:1447973
A;Accession: S25273
A;Molecule type: DNA
A;Residues: 1-75 <DIS>
A;Cross-references: EMBL:X04449; NID:g15549; PID:g15550

Query Match          95.0%; Score 19; DB 2; Length 75;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

A;Molecule type: DNA
A;Residues: 1-64 <KVI>
A;Cross-references: GB:L26909; NID:G598395; PIDN:AA42221.1; PID:G598396
C;Genetics:
A;Gene: Tf
A;Introns: 14/1
C;Superfamily: transferrin; transferrin repeat homology

Query Match          95.0%; Score 19; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 VAEF 5
DB      49 VAEF 52

RESULT 13
T18034
hypothetical protein A532L - Chlorella virus PBCV-1
C;Species: Chlorella virus PBCV-1
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18034
R;Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A;Reference number: Z18806
A;Accession: T18034
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-79 <GRA>
A;Cross-references: EMBL:U42580; NID:94028896; PIDN:AA096899.1
A;Experimental source: specific host Chlorella strain NC64A
C;Genetics:
A;Note: A532L

Query Match          95.0%; Score 19; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 VAEF 5
DB      49 VAEF 52

RESULT 14
T27605
hypothetical protein ZC477.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27605
R;Du, Z.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid ZC477.
A;Reference number: Z20392
A;Accession: T27605
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-79 <DUZ>
A;Cross-references: EMBL:U40802; PIDN:AAA81506.1; CBSP:ZC477.6
C;Genetics:
A;Gene: CBSP:ZC477.6
A;Introns: 51/2

Query Match          95.0%; Score 19; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 VAEF 5
DB      74 VAEF 77

RESULT 15
F83922
hypothetical protein BH2182 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: F83922
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir,
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: F83922
A;Status: preliminary
A;Molecule type: DNA

```

A;Cross-residues: 1-84 <STO>  
 A;Cross-references: GB:AF001514; GB:BA000004; NID:gl0174613; PIDN:BA05901.1; GSPDB:GN00  
 A;Experimental source: strain C-125  
 C;Genetics:

A;Gene: BH2182

Query Match 95.0%; Score 19; DB 2; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 38 VAEF 41

#### RESULT 16

A;Cross-residues: 1-84 <STO>  
 A;Cross-references: GB:AF001514; GB:BA000004; NID:gl0174613; PIDN:BA05901.1; GSPDB:GN00  
 A;Experimental source: strain C-125  
 C;Genetics:

Query Match 95.0%; Score 19; DB 2; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 13 VAEF 16

#### RESULT 17

A;Cross-residues: 1-84 <STO>  
 A;Cross-references: GB:AF001514; GB:BA000004; NID:gl0174613; PIDN:BA05901.1; GSPDB:GN00  
 A;Experimental source: strain C-125  
 C;Genetics:

Query Match 95.0%; Score 19; DB 2; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 38 VAEF 41

#### RESULT 18

A;Cross-residues: 1-84 <STO>  
 A;Cross-references: GB:AF001514; GB:BA000004; NID:gl0174613; PIDN:BA05901.1; GSPDB:GN00  
 A;Experimental source: strain C-125  
 C;Genetics:

Query Match 95.0%; Score 19; DB 2; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 38 VAEF 41

Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 12 VAEF 15

QY 2 VAEF 5  
 DB 68 VAEF 71

#### RESULT 19

A;Cross-residues: 1-84 <STO>  
 A;Cross-references: GB:AF001514; GB:BA000004; NID:gl0174613; PIDN:BA05901.1; GSPDB:GN00  
 A;Experimental source: strain C-125  
 C;Genetics:

QY 2 VAEF 5  
 DB 3 VAEF 6

QY 2 VAEF 5  
 DB 3 VAEF 6

QY 2 VAEF 5  
 DB 3 VAEF 6

#### RESULT 20

A;Cross-residues: 1-84 <STO>  
 A;Cross-references: GB:AF001514; GB:BA000004; NID:gl0174613; PIDN:BA05901.1; GSPDB:GN00  
 A;Experimental source: strain C-125  
 C;Genetics:

QY 2 VAEF 5  
 DB 3 VAEF 6

QY 2 VAEF 5  
 DB 3 VAEF 6

Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AB3053  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-96 <KUR>  
 A:Cross-references: GB:AE008689; PIDN:AAI44840.1; PID:g17742484; GSPDB:GN00187  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atu4039  
 A:Map position: linear chromosome

Query Match 95.0%; Score 19; DB 2; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 Db 55 VAEF 58

RESULT 21  
 AC1397  
 conserved hypothetical protein lmo2579 [imported] - Listeria monocytogenes (strain EGD-e  
 C:Species: Listeria monocytogenes  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AC1397  
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bleecker  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kuraptat, G.; Madueno, E.; Maitournam, A.; Ma  
 ok, C.; Schlucker, T.; Simoes, N.; Trierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
 A:Title: Comparative Genomics of Listeria species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AC1397  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-97 <GLA>  
 A:Cross-references: GB:NC\_003210; PIDN:CAD00657.1; PID:g16412067; GSPDB:GN00177  
 A:Experimental source: strain EGD-e  
 C:Genetics:  
 A:Gene: lmo2579

Query Match 95.0%; Score 19; DB 2; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 Db 15 VAEF 18

RESULT 22  
 E83244  
 conserved hypothetical protein PA3202 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: E83244  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: AB2950; MUID:20437337; PMID:10984043  
 A:Accession: E83244  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-99 <STO>  
 A:Cross-references: GB:AE004744; GB:AE004091; NID:g9949317; PIDN:AAG06590.1; GSPDB:GN001

A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA3202

Query Match 95.0%; Score 19; DB 2; Length 99;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 Db 62 VAEF 65

## RESULT 23

H90297

conserved hypothetical protein [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001

C:Accession: H90297

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,  
 arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.  
 submitted to Genbank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: H90297

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-101 &lt;KUR&gt;

A:Cross-references: GB:AE006641; NID:g13814625; PIDN:AAK41639.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO1404

Query Match 95.0%; Score 19; DB 2; Length 101;

Best Local Similarity 100.0%; Pred. No. 2.8e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 Db 20 VAEF 23

## RESULT 24

Tl0097

nifX protein - Methanococcus maripaludis

C:Species: Methanococcus maripaludis

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000

C:Accession: Tl0097

R:Kessler, P.S.; McLarnan, J.; Leigh, J.A.

J. Bacteriol. 179, 541-543, 1997

A:Title: Nitrogenase phylogeny and the molybdenum dependence of nitrogen fixation in Me

A:Reference number: Z16944; MUID:97144542; PMID:8990309

A:Accession: Tl0097

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-106 &lt;XES&gt;

A:Cross-references: EMBL:U75887; NID:g1666882; PIDN:AA045519.1; PID:g1666890

A:Experimental source: strain JJ

C:Genetics:

A:Gene: nifX

Query Match 95.0%; Score 19; DB 2; Length 106;

Best Local Similarity 100.0%; Pred. No. 3e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 Db 35 VAEF 38

## RESULT 25

C64490

hypothetical protein M1524 - Methanococcus jannaschii

```

;Species: Methanococcus jannaschii
;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 12-Jun-2003
;Accession: C64490
;Buit, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodok, A.;
;son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hirst, M.A.
;Science 273, 1058-1073, 1996
;Authors: Kaine, B.F.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
;Reference number: A64300; MUID:96337999; PMID:8688087
;Accession: C64490
;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-108 <BUL>
;Cross-references: GB:U67593; GB:L77117; NID:G2826427; PIDN:AA99550.1; PID:G1592154; T
;Genetics:
;Map position: FOR1501406-1501732
;Superfamily: uncharacterized conserved protein MJ1524

Query Match 95.0%; Score 19; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
DB 52 VAEF 55

RESULT 26
E70949
Probable regulator - Mycobacterium tuberculosis (strain H37Rv)
;Species: Mycobacterium tuberculosis
;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
;Accession: E70949
;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
; Nature 393, 537-544, 1998
;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
;Reference number: A70500; MUID:98295987; PMID:9634230
;Accession: E70949
;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-109 <COL>
;Cross-references: GB:AL021646; GB:AL123456; NID:G3242278; PIDN:CAA16648.1; PID:G282759
;Experimental source: strain H37Rv
;Genetics:
;Gene: RV3183

Query Match 95.0%; Score 19; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
DB 97 VAEF 100

RESULT 27
T17582
Hypothetical protein A92L - Chlorella virus PBCV-1
;Species: Chlorella virus PBCV-1
;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
;Accession: T17582
;Graves, M.V.; Van Etten, J.L.
;Submitted to the EMBL Data Library, May 1999
;Reference number: Z18806
;Accession: T17582
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-111 <GRA>
;Cross-references: EMBL:U42580; NID:G4028896; PIDN:ARC96460.1
;Experimental source: specific host Chlorella strain NC64
;Genetics:
;Gene: A92L
;Superfamily: Chlorella virus PBCV-1 hypothetical protein A92L

Query Match 95.0%; Score 19; DB 2; Length 111;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
DB 17 VAEF 20

RESULT 28
B83431
type III export protein PscI PA1722 [imported] - Pseudomonas aeruginosa (strain PA01)
;Species: Pseudomonas aeruginosa
;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
;Accession: B83431
;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Br
; adman, S.; Yuan, Y.; Brody, L.B.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
; Nature 406, 959-964, 2000
;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
; Reference number: A82950; MUID:20437337; PMID:10984043
;Accession: B83431
;Status: preliminary
;Molecule type: DNA
;Residues: 1-112 <STO>
;Cross-references: GB:AE004598; GB:AE004091; NID:G9947687; PIDN:AAG05111.1; GSPDB:GN001
;Experimental source: strain PA01
;Genetics:
;Gene: pscI; PA1722
;Superfamily: Yersinia enterocolitica plasmid pYV virC-region hypothetical protein yscI

Query Match 95.0%; Score 19; DB 2; Length 112;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
DB 28 VAEF 31

RESULT 29
A55872
Kedarcidin [validated] - Streptococcus sp. (ATCC 53650)
;Alternate names: Kedarcidin apoprotein
;Species: Streptococcus sp.
;Variety: strain L585-6 (ATCC 53650)
;Date: 21-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 15-Sep-2000
;Accession: A55872
;Hofstead, S.J.; Watson, J.A.; Malacko, A.R.; Marquardt, H.
; J. Antibiot. 45, 1250-1254, 1992
;Title: Kedarcidin, a new chromoprotein antitumor antibiotic. II. Isolation, purificati
; Reference number: A55872; MUID:93015257; PMID:1399845
;Accession: A55872
;Molecule type: protein
;Residues: 1-114 <HOF>
;Note: sequence extracted from NCBI backbone (NCBI:P118354)
;Constantine, K.L.; Colson, K.L.; Wittekind, M.; Friedrichs, M.S.; Zein, N.; Tuttle, J.
; ler, L.
; submitted to the Brookhaven Protein Data Bank, June 1994
;Reference number: A52519; PDB:1AKP
;Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR, residue
;Zein, N.; Casazza, A.M.; Doyle, T.W.; Leet, J.E.; Schroeder, D.R.; Solomon, W.; Nadler
; Proc. Natl. Acad. Sci. U.S.A. 90, 8009-8012, 1993
;Title: Selective proteolytic activity of the antitumor agent kedarcidin.
;Reference number: A58601; MUID:9336732; PMID:8367457
;Contents: annotation; protein activity
;Note: the apoprotein may have proteolytic activity
;Superfamily: macromycin

```

C;Keywords: antibiotic; pigment binding  
F;37-47,88-95/Disulfide bonds: #status experimental

Query Match 95.0%; Score 19; DB 2; Length 114;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
Db 49 VAEF 52

RESULT 30  
D97414  
hypothetical protein AGR C 804 [imported] - Agrobacterium tumefaciens (strain C58, Cerec  
C;Species: Agrobacterium tumefaciens  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C;Accession: D97414  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: D97414  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-114 <KUR>  
A;Cross-references: GB:AE007869; PIDN:AAK86269.1; PID:g15155379; GSPDB:GN00169  
C;Genetics:  
A;Gene: AGR C 804  
A;Map position: circular chromosome

Query Match 95.0%; Score 19; DB 2; Length 114;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
Db 40 VAEF 43

RESULT 31  
D75218  
hypothetical protein PRB0185 - Pyrococcus abyssi (strain Orsay)  
C;Species: Pyrococcus abyssi  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C;Accession: D75218  
R;anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru  
A;Reference number: A75001  
A;Accession: D75218  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-115 <RAW>  
A;Cross-references: GB:AJ248283; GB:AL096836; NID:G5457433; PIDN:CAB49195.1; PID:e151508  
A;Experimental source: strain Orsay  
C;Genetics:  
A;Gene: PRB0185

Query Match 95.0%; Score 19; DB 2; Length 115;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
Db 101 VAEF 104

RESULT 32  
IPAF  
insulin precursor - American goosfish  
C;Species: Lophius americanus (American goosfish)

C;Date: 31-Mar-1981 #sequence\_revision 31-Mar-1981 #text\_change 16-Jul-1999  
C;Accession: A01608  
R;Hobart, P.M.; Shen, L.P.; Crawford, R.; Pictet, R.L.; Rutter, W.J.  
Science 210, 1360-1363, 1980  
A;Title: Comparison of the nucleic acid sequence of anglerfish and mammalian insulin mR  
A;Reference number: A94254; MUID:81056434; PMID:7001633  
A;Accession: A01608  
A;Molecule type: mRNA  
A;Residues: 1-116 <HOB>  
A;Cross-references: GB:V00634; GB:J00934; NID:G64025; PIDN:CAA23907.1; PID:G64026  
C;Superfamily: insulin  
C;Keywords: hormone, pancreas  
F;1-24/Domain: signal sequence #status predicted <SIG>  
F;25-54/Domain: insulin chain B #status predicted <BC>  
F;55-54,96-116/Product: insulin #status predicted <MAT>  
F;56-93/Domain: connecting peptide #status predicted <CP>  
F;96-116/Domain: insulin chain A #status predicted <ACH>  
F;32-102,44-115,101-106/Disulfide bonds: #status predicted

Query Match 95.0%; Score 19; DB 1; Length 116;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
Db 80 VAEF 83

RESULT 33  
T48682  
hypothetical protein DKFZp761N05121.1 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 05-May-2000  
C;Accession: T48682  
R;Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, April 2000.  
A;Reference number: Z24535  
A;Accession: T48682  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-118 <AAA>  
A;Cross-references: EMBL:AL353932  
A;Experimental source: adult amygdala; clone DKFZp761N05121  
C;Genetics:  
A;Note: DKFZp761N05121.1

Query Match 95.0%; Score 19; DB 2; Length 118;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
Db 15 VAEF 18

RESULT 34  
F95327  
hypothetical protein SMO974 [imported] - Sinorhizobium meliloti (strain 1021) magaplas  
C;Species: Sinorhizobium meliloti  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C;Accession: F95327  
R;Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bow  
; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yen, K.C  
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilo  
A;Reference number: A95262; MUID:21396509; PMID:11481432  
A;Accession: F95327  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-119 <KUR>  
A;Cross-references: GB:AE006469; PIDN:AAK65184.1; PID:g14523629; GSPDB:GN00165  
A;Experimental source: strain 1021, megaplasmid pSymA  
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, F.; Ampe, F.; Barloy-Hubler

ela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Kelaure, J.; Leubolt, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 A:Genetics:  
 A:Gene: Sma0974  
 A:Genome: plasmid

Query Match 95.0%; Score 19; DB 2; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 Db 26 VAEF 29

RESULT 35  
 782861  
 conjugal transfer protein Xfa0006 [imported] - Xylella fastidiosa (strain 9a5c)  
 A:Species: Xylella fastidiosa  
 A:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 A:Accession: F82861  
 A:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: F82861  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-122 <STM>  
 A:Cross-references: GB:AE003851; NID:g9112238; PIDN:AAF85575.1; GSPDB:GN00130; XFSC:XFA00006  
 A:Experimental source: strain 9a5c  
 A:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Azeiteiro, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, H. S.; Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira A.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328  
 A:Contents: annotation  
 A:Genetics:  
 A:Gene: Xfa0006  
 A:Genome: plasmid  
 A:Note: plasmid pXF5.1

Query Match 95.0%; Score 19; DB 2; Length 122;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 Db 32 VAEF 35

RESULT 36  
 369129  
 gamma-carboxymuconolactone decarboxylase - Methanobacterium thermoautotrophicum (strain gamma-129)  
 A:Species: Methanobacterium thermoautotrophicum  
 A:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Jun-1999  
 A:Accession: B69129  
 A:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;

Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N. K.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997  
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct  
 A:Reference number: A69000; MUID:98037514; PMID:9371463  
 A:Accession: B69129  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-125 <MTH>  
 A:Cross-references: GB:AE000810; GB:AE000666; NID:g2621277; PIDN:AAB84740.1; PID:g262126  
 A:Experimental source: strain Delta H  
 A:Genetics:  
 A:Gene: Mtn234  
 A:Superfamily: 4-carboxymuconolactone decarboxylase

Query Match 95.0%; Score 19; DB 2; Length 125;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 Db 37 VAEF 40

RESULT 37  
 D70790  
 Hypothetical protein Rv3675 - Mycobacterium tuberculosis (strain H37RV)  
 A:Species: Mycobacterium tuberculosis  
 A:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 A:Accession: D70790  
 A:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S. Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S. Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: D70790  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-125 <COL>  
 A:Cross-references: GB:AL022121; GB:AL123456; NID:g3261559; PIDN:CAA17997.1; PID:el26453  
 A:Experimental source: strain H37RV  
 A:Genetics:  
 A:Gene: Rv3675

Query Match 95.0%; Score 19; DB 2; Length 125;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 Db 42 VAEF 45

RESULT 38  
 B81368  
 Hypothetical protein Cj0939c [imported] - Campylobacter jejuni (strain NCTC 11168)  
 A:Species: Campylobacter jejuni  
 A:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 A:Accession: B81368  
 A:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell Nature 403, 665-668, 2000  
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp A:Reference number: A81250; MUID:20150912; PMID:10686204  
 A:Accession: B81368  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-126 <PAR>  
 A:Cross-references: GB:AL139076; GB:AL111168; NID:g6968128; PIDN:CAB73196.1; PID:g696837  
 A:Experimental source: serotype O2, strain NCTC 11168  
 A:Genetics:

A;Gene: Cj0938c

Query Match 95.0%; Score 19; DB 2; Length 126;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 Db 83 VAEF 86

## RESULT 39

D70799 hypothetical protein Rv3747 - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis  
 C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C;Accession: D70799  
 R;Cole, S.T.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Bresch, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998  
 A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.  
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: D70799  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA

A;Residues: 1-127 <COL>  
 A;Cross-references: GB:AL022121; GB:AL123456; NID:G3261559; PIDN:CAA18069.1; PID:G296017  
 A;Experimental source: strain H37RV  
 C;Genetics:  
 A;Gene: Rv3747

Query Match 95.0%; Score 19; DB 2; Length 127;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 Db 80 VAEF 83

## RESULT 40

AB1732 hypothetical protein lin2399 [imported] - Listeria innocua (strain Clip11262)

C;Species: Listeria innocua  
 C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C;Accession: AB1732  
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feigl, H.; Jones, L.N.; Karst, U.

Science 294, 849-852, 2001  
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative Genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AB1732  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-127 <GLA>  
 A;Cross-references: GB:AL592022; PIDN:CAC97626.1; PID:G16414922; GSPDB:GN00178  
 A;Experimental source: strain Clip11262  
 C;Genetics:  
 A;Gene: lin2399

Query Match 95.0%; Score 19; DB 2; Length 127;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 ||||  
 Db 71 VAEF 74

## RESULT 41

A54797 ileal lipid-binding protein - mouse  
 C;Species: Mus musculus (house mouse)

C;Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 05-Nov-1999  
 C;Accession: A54797

R;Crossman, M.W.; Haft, S.M.; Gordon, J.I.

J. Cell Biol. 126, 1547-1564, 1994

A;Title: The mouse ileal lipid-binding protein gene: a model for studying axial pattern

A;Reference number: A54797; MUID:94375529; PMID:8089185

A;Accession: A54797

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-128 <CRO>

A;Cross-references: GB:U00938; NID:G507143; PIDN:AAC27352.1; PID:G507144

C;Genetics:

A;Introns: 23/1; 81/3; 111/3

C;Superfamily: myelin P2 protein

Query Match 95.0%; Score 19; DB 2; Length 128;

Best Local Similarity 100.0%; Pred. No. 3.6e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5

||||

Db 92 VAEF 95

## RESULT 42

T02012

probable cathepsin B-like cysteine proteinase (EC 3.4.22.-) T15B16.17b - Arabidopsis th  
 C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 22-Jun-1999

C;Accession: T02012

R;Stonking, T.; Smith, R.

submitted to the EMBL Data Library, November 1998

A;Description: The sequence of A. thaliana T15B16.

A;Reference number: Z14488

A;Accession: T02012

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-129 <STO>

A;Cross-references: EMBL:AF104919; NID:G3859590; PIDN:AAC72873.1; PID:G3859607

A;Experimental source: cultivar Columbia

C;Genetics:

A;Map position: 4; 44/3; 71/3

A;Introns: 26/3; 44/3; 71/3

A;Note: T15B16.17b

C;Superfamily: papain

C;Keywords: cysteine proteinase; hydrolase

Query Match 95.0%; Score 19; DB 2; Length 129;

Best Local Similarity 100.0%; Pred. No. 3.7e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5

||||

Db 72 VAEF 75

## RESULT 43

D72348

conserved hypothetical protein - Thermotoga maritima (strain MSB8)

C;Species: Thermotoga maritima

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C;Accession: D72348

R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hicke

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.

C.M.

Nature 395, 323-329, 1999

A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome se

A;Reference number: A72200; MUID:99287316; PMID:10360571

```

A:Accession: D72348
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-130 <ARN>
A:Cross-references: GB:AE001739; GB:AE000512; NID:94981176; PID:9498117
C:Genetics:
C:Superfamily: conserved hypothetical protein W0315
Query Match 95.0%; Score 19; DB 2; Length 130;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 VAEF 5
DB 9 VAEF 12

RESULT 44
375208
hypothetical protein PAB2224 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 06-Jan-2003
C:Accession: G75208
R:anonymous, Genoscope
Submitted to the ENBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: G75208
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-131 <KAW>
A:Cross-references: GB:AJ249283; GB:AL096836; NID:95457433; PIDN:CAB49118.1; PID:9545762
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB2224
C:Superfamily: uncharacterized conserved protein
Query Match 95.0%; Score 19; DB 2; Length 131;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 VAEF 5
DB 4 VAEF 7

RESULT 45
JQ2270
hypothetical 15.2K protein - Anabaena sp.
C:Species: Anabaena sp.
C>Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 13-Sep-1998
C:Accession: JQ2270
R:Raimier, F.W.; Soper, T.S.
Gene 126, 85-92, 1993
A>Title: Overproduction of Anabaena 7120 ribulose-bisphosphate carboxylase/oxygenase in
A:Reference number: JQ2270; MUID:93231541; PMID:8472962
A:Accession: JQ2270
A:Molecule type: DNA
A:Residues: 1-132 <LAR>
A:Cross-references: DDBJ:J01540
A:Experimental source: strain 7120
C:Genetics:
A:Gene: rbcX
C:Superfamily: rbcX protein
Query Match 95.0%; Score 19; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 VAEF 5
DB 9 VAEF 12

RESULT 46
AG1996
hypothetical protein alr1525 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1996
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi:
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, I.
DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1996
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-132 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA077891.1; PID:G17135345; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1525
C:Superfamily: rbcX protein
Query Match 95.0%; Score 19; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 VAEF 5
DB 81 VAEF 84

RESULT 47
AH0137
conserved hypothetical protein YP01120 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AH0137
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
ii, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AH0137
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-133 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89963.1; PID:G15979187; GSPDB:GN00175
C:Genetics:
A:Gene: YP01120
C:Superfamily: 15.5K protein (tolAB operon 5' region)
Query Match 95.0%; Score 19; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 VAEF 5
DB 129 VAEF 132

RESULT 48
H75623
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: H75623
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

```

Science 286, 1571-1577, 1999  
 A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036996; PMID:10567266  
 A:Accession: H75623  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-133 <WHI>  
 A:Cross-references: GB:AE001826; NID:g6460827; PIDN:AAF12590.1; PID:g6460886; TIGR:DRB00  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DRB0052  
 A:Map position: megaplasmid  
 A:Genome: plasmid  
 A:Note: plasmid MP1

Query Match 95.0%; Score 19; DB 2; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 Db 18 VAEF 21

RESULT 49  
 A84221  
 riboflavin synthase beta subunit [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: A84221  
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;  
 ; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
 A>Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483; PMID:11016950  
 A:Accession: A84221  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-133 <STO>  
 A:Cross-references: GB:AE004437; NID:g10580221; PIDN:AAG19133.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: ribB

Query Match 95.0%; Score 19; DB 2; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 Db 8 VAEF 11

RESULT 50  
 WNEC15  
 ybgC protein - Escherichia coli (strain K-12)  
 C:Species: Escherichia coli  
 C>Date: 15-Oct-1982 #sequence\_revision 30-Jun-1988 #text\_change 01-Mar-2002  
 C:Accession: A25980; G64809  
 R:Sun, T.P.; Webster, R.E.  
 J. Bacteriol. 169, 2667-2674, 1987  
 A>Title: Nucleotide sequence of a gene cluster involved in entry of E colicins and singl  
 A:Reference number: A91835; MUID:97222192; PMID:3294803  
 A:Accession: A25980  
 A:Molecule type: DNA  
 A:Residues: 1-134 <SUN>  
 A:Cross-references: GB:M16489; NID:g148021; PIDN:AAAB3918.1; PID:g1128977  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A>Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64809  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-134 <BLAT>  
 A:Cross-references: GB:AE000177; GB:U00096; NID:g1786955; PIDN:AACT73830.1; PID:g1786957  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Comment: This is one of the proteins, encoded by the *fil*-*tol*AB gene cluster, that is  
 C:Genetics:  
 A:Gene: ybgC  
 A:Map position: 17 min  
 A:Start codon: GTG  
 C:Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match 95.0%; Score 19; DB 1; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 Db 129 VAEF 132

RESULT 51  
 C90725  
 hypothetical protein ECs0771 [imported] - Escherichia coli (strain O157:H7, substrain F  
 C:Species: Escherichia coli  
 C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C:Accession: C90725  
 R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.C  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
 A:Reference number: A99625; MUID:21156231; PMID:11258796  
 A:Accession: C90725  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-134 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA334194.1; PID:g13360230; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: ECs0771  
 C:Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match 95.0%; Score 19; DB 2; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VAEF 5  
 Db 129 VAEF 132

RESULT 52  
 D85576  
 hypothetical protein ybgC [imported] - Escherichia coli (strain O157:H7, substrain EDL9  
 C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: D85576  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayne  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dialanta, E.; Potamouais, K.; Apodaca  
 Nature 409, 529-533, 2001  
 A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: D85576  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-134 <STO>  
 A:Cross-references: GB:AE005174; NID:g12513669; PIDN:AAAS5072.1; GSPDB:GN00145; UWGP:Z0  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: ybgC  
 C:Superfamily: 15.5K protein (tolAB operon 5' region)

```

Query Match      95.0%; Score 19; DB 2; Length 134;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
  ||||
b 129 VAEF 132

RESULT 53
CD0592
conserved hypothetical protein STX0790 [imported] - Salmonella enterica subsp. enterica
;Species: Salmonella enterica subsp. enterica serovar Typhi
;Note: this species has also been called Salmonella typhi
;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
;Accession: AD0592
;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
;T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
;S.; Moutle, S.; O'Gaora, P.
ature 413, 846-852, 2001
;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
;Reference number: AB0502; MUID:21534947; PMID:11677608
;Accession: AD0592
;Status: preliminary
;Molecule type: DNA
;Residues: 1-134 <PAR>
;Cross-references: GB:AL513382; PIDN:CAD05206.1; PID:gl6501976; GSPDB:GN00176
;Genetics:
;Gene: STX0790
;Superfamily: 15.5K protein (tolAB operon 5' region)

Query Match      95.0%; Score 19; DB 2; Length 134;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
  ||||
b 129 VAEF 132

RESULT 54
;76416
hypothetical protein - Synecocystis sp. (strain PCC 6803)
;Species: Synecocystis sp.
;Variety: PCC 6803
;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
;Accession: S76416
;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
;K. Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
;DNA Res. 3, 109-136, 1996
;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
;Reference number: S74322; MUID:97061201; PMID:8905231
;Accession: S76416
;Status: preliminary
;Molecule type: DNA
;Residues: 1-136 <KAN>
;Cross-references: EMBL:D90915; GB:AB001339; NID:gl653604; PIDN:BAAL18545.1; PID:gl65363
;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match      95.0%; Score 19; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
  ||||
b 88 VAEF 91

RESULT 55
E81355
hypothetical protein Cj0829c [imported] - Campylobacter jejuni (strain NCTC 11168)

```

```

C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: E81355
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hy
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: E81355
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-137 <PAR>
A;Cross-references: GB:AL139076; GB:AL111168; NID:G6968128; PIDN:CAB73094.1; PID:G696827
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
C;Gene: Cj0829c
C;Superfamily: hypothetical protein ynfF

Query Match      95.0%; Score 19; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
  ||||
b 33 VAEF 36

RESULT 56
F82106
flagellar basal-body rod protein FlgC VC2199 [imported] - Vibrio cholerae (strain N16961
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: F82106
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermlaeva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragoi, I.; Sellers, I
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: F82106
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-138 <HEI>
A;Cross-references: GB:AE004292; GB:AE003852; NID:G956760; PIDN:AAF95344.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
C;Gene: VC2199
A;Map position: 1

Query Match      95.0%; Score 19; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
  ||||
b 79 VAEF 82

RESULT 57
A69052
riboflavin synthase beta subunit - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-May-2000
C;Accession: A69052
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: A69052
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

```

A;Residues: 1-139 <MTH>  
A;Cross-references: GB:AE000902; GB:AE000666; NID:g2622500; PIDN:AAB85867.1; PID:g262250  
A;Experimental source: strain Delta H  
C;Genetics:  
A;Gene: MHL390  
C;Superfamily: riboflavin synthase beta chain

Query Match 95.0%; Score 19; DB 2; Length 139;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
|||  
Db 10 VAEF 13

RESULT 58  
A83975  
hypothetical protein BH2601 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: A83975  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A;Reference number: A83650; MUID:20512582; PMID:11059132  
A;Accession: A83975  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-139 <STO>  
A;Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA06320.1; GSPDB:GN00  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH2601  
C;Superfamily: Bacillus subtilis hypothetical protein yIbD

Query Match 95.0%; Score 19; DB 2; Length 139;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
|||  
Db 11 VAEF 14

RESULT 59  
B69323  
hypothetical protein AF0586 - Archaeoglobus fulgidus  
C;Species: Archaeoglobus fulgidus  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Nov-1999  
C;Accession: B69323  
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A;Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae  
A;Reference number: A69250; MUID:98049343; PMID:9389475  
A;Accession: B69323  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-140 <KLE>  
A;Cross-references: GB:AE001064; GB:AE000782; NID:g2689387; PIDN:AAB90658.1; PID:g265004  
C;Superfamily: Archaeoglobus fulgidus hypothetical protein AF0586

Query Match 95.0%; Score 19; DB 2; Length 140;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
|||  
Db 65 VAEF 68

## RESULT 60

JQ2279

low-temperature regulated protein BN115 - rape

C;Species: Brassica napus (rape)

C;Date: 30-Sep-1993 #sequence\_revision 20-Aug-1994 #text\_change 28-May-1999

C;Accession: JQ2279

R;Weretilnyk, E.; Orr, W.; White, T.C.; Iu, B.; Singh, J.

Plant Physiol. 101, 171-177, 1993

A;Title: Characterization of three related low-temperature-regulated cDNAs from winter

A;Reference number: JQ2279; MUID:94105287; PMID:7904076

A;Accession: JQ2279

A;Molecule type: mRNA

A;Residues: 1-142 &lt;WER&gt;

A;Cross-references: GB:S66726; NID:g544693; PIDN:AAB29482.1; PID:g544694

C;Comment: This protein is a low-temperature-related protein.

Query Match 95.0%; Score 19; DB 2; Length 142;

Best Local Similarity 100.0%; Pred. No. 4e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5

|||

Db 125 VAEF 128

## RESULT 61

JQ2280

low-temperature regulated protein BN19 - rape

C;Species: Brassica napus (rape)

C;Date: 30-Sep-1993 #sequence\_revision 20-Aug-1994 #text\_change 28-May-1999

C;Accession: JQ2280

R;Weretilnyk, E.; Orr, W.; White, T.C.; Iu, B.; Singh, J.

Plant Physiol. 101, 171-177, 1993

A;Title: Characterization of three related low-temperature-regulated cDNAs from winter

A;Reference number: JQ2279; MUID:94105287; PMID:7904076

A;Accession: JQ2280

A;Molecule type: mRNA

A;Residues: 1-142 &lt;WER&gt;

A;Cross-references: GB:S68879; NID:g544695; PIDN:AAB29483.1; PID:g544696

C;Comment: This protein is a low-temperature-related protein.

Query Match 95.0%; Score 19; DB 2; Length 142;

Best Local Similarity 100.0%; Pred. No. 4e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5

|||

Db 125 VAEF 128

## RESULT 62

H69515

riboflavin synthase, subunit beta (ribs) homolog - Archaeoglobus fulgidus

C;Species: Archaeoglobus fulgidus

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jul-2000

C;Accession: H69515

R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodso

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F

; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A;Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S

Smith, H.O.; Woese, C.R.; Venter, J.C.

A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A;Reference number: A69250; MUID:98049343; PMID:9389475

A;Accession: H69515

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-143 &lt;KLE&gt;

A;Cross-references: GB:AE000957; GB:AE000782; NID:g2689280; PIDN:AAB89124.1; PID:g26484

C;Superfamily: riboflavin synthase beta chain

Query Match 95.0%; Score 19; DB 2; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 4.1e+02; Indels 0; Gaps 0;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

iy 2 VAEF 5  
 ||||  
 10 VAEF 13

RESULT 63  
 36978  
 probable transposase, truncated section 1 [imported] - Streptomyces coelicolor (fragment)  
 ;Species: Streptomyces coelicolor  
 ;Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
 ;Accession: T36978  
 ;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Bartell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, August 1999  
 ;Reference number: Z21618  
 ;Accession: T36978  
 ;Status: preliminary; translated from GB/EMBL/DDBJ  
 ;Molecule type: DNA  
 ;Residues: 1-143 <OLI>  
 ;Cross-references: EMBL:AL109849; PIDN:CAB52892.1; GSPDB:GN00070; SCOEDB:SCJ11.07c  
 ;Experimental source: strain A3(2)  
 ;Comment: A complete transposase sequence can be assembled from three adjacent orfs (PI  
 ;Genetics:  
 ;Gene: SCOEDB:SCJ11.07c

Query Match 95.0%; Score 19; DB 2; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 4.1e+02; Indels 0; Gaps 0;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

iy 2 VAEF 5  
 ||||  
 113 VAEF 116

RESULT 64  
 341715  
 ribosomal protein L13 [similarity] - Haloarcula marismortui  
 ;Species: Haloarcula marismortui  
 ;Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 14-Apr-2003  
 ;Accession: B41715  
 ;Kroemer, W.J.; Arndt, B.  
 J. Biol. Chem. 266, 24573-24579, 1991  
 ;Title: Halobacterial S9 operon. Three ribosomal protein genes are cotranscribed with 9  
 bacterium) marismortui.  
 ;Reference number: A41715; MUID:92105119; PMID:1840597  
 ;Accession: B41715  
 ;Status: preliminary  
 ;Molecule type: DNA  
 ;Residues: 1-145 <KRO>  
 ;Cross-references: GB:M76567; NID:gl48775; PIDN:AAA73097.1; PID:gl48777  
 ;Superfamily: ribosomal protein L13

Query Match 95.0%; Score 19; DB 1; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 4.1e+02; Indels 0; Gaps 0;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

iy 2 VAEF 5  
 ||||  
 3 VAEF 6

RESULT 65  
 S77043  
 hypothetical protein sll0678 - Synecocystis sp. (strain PCC 6803)  
 ;Species: Synecocystis sp.  
 ;Variety: PCC 6803  
 ;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
 ;Accession: S77043  
 ;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996  
 A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis  
 S.  
 A;Reference number: S74322; MUID:97061201; PMID:8905231  
 A;Accession: S77043  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-145 <KAN>  
 A;Cross-references: EMBL:D64005; GB:AB001339; NID:gl001779; PIDN:BAAL0735.1; PID:gl00651  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C;Superfamily: Synecocystis hypothetical protein sll0678

Query Match 95.0%; Score 19; DB 2; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 4.1e+02; Indels 0; Gaps 0;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

iy 2 VAEF 5  
 ||||  
 35 VAEF 38

RESULT 66  
 F89837  
 conserved hypothetical protein SA0624 [imported] - Staphylococcus aureus (strain N315)  
 C;Species: Staphylococcus aureus  
 C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C;Accession: F89837  
 C;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguni, A.; Mitani, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
 Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001

A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A;Reference number: A89758; MUID:21311952; PMID:11418146  
 A;Accession: F89837  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-146 <KUR>  
 A;Cross-references: GB:BA000018; PID:gl3700560; PIDN:BAE41857.1; GSPDB:GN00149  
 A;Experimental source: strain N315  
 C;Genetics:  
 C;Gene: SA0624

Query Match 95.0%; Score 19; DB 2; Length 146;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02; Indels 0; Gaps 0;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

iy 2 VAEF 5  
 ||||  
 102 VAEF 105

RESULT 67  
 C94294  
 hypothetical protein Vngl398c [imported] - Halobacterium sp. NRC-1  
 C;Species: Halobacterium sp. NRC-1  
 C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
 C;Accession: C94294  
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, E.  
 ;Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
 A>Title: Genome sequence of Halobacterium species NRC-1.  
 A;Reference number: A84160; MUID:20504483; PMID:11016950  
 A;Accession: C94294  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-147 <STO>  
 A;Cross-references: GB:AE004437; NID:gl0580903; PIDN:AAG19719.1; GSPDB:GN00138  
 C;Genetics:  
 C;Gene: VNGl398c  
 C;Superfamily: hypothetical protein yJcf

Query Match 95.0%; Score 19; DB 2; Length 147;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 |||||  
 DB 117 VAEF 120

RESULT 68  
 T34803  
 hypothetical protein SC2E1.35 - Streptomyces coelicolor  
 C/Species: Streptomyces coelicolor  
 C/Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 04-Mar-2000  
 C/Accession: T34803  
 R/Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, June 1998  
 A/Reference number: Z21557  
 A/Accession: T34803  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-147 <MUR>  
 A/Cross-references: EMBL:AL023797; PIDN:CAA19410.1; GSPDB:GN00070; SCODEB:SC2E1.35  
 A/Experimental source: strain A3(2)  
 C/Genetics:  
 A/Gene: SCODEB:SC2E1.35  
 C/Superfamily: Streptomyces coelicolor hypothetical protein SC2E1.35

Query Match 95.0%; Score 19; DB 2; Length 147;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 |||||  
 DB 125 VAEF 128

RESULT 69  
 E81788  
 conserved hypothetical protein NMA2160 [imported] - Neisseria meningitidis (strain Z2491)  
 C/Species: Neisseria meningitidis  
 C/Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
 C/Accession: E81788  
 R/Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
 ; Holroyd, S.; Jagels, K.; Leather, S.; Mungall, K.; Quail, M.A.; Rajandream,  
 Nature 404, 502-506, 2000  
 A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
 A/Reference number: A81775; MUID:20222536; PMID:10761919  
 A/Accession: E81788  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-148 <PAR>  
 A/Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85372.1; PID:g738078  
 A/Experimental source: serogroup A, strain Z2491  
 C/Genetics:  
 A/Gene: NMA2160  
 C/Superfamily: hypothetical protein ytwi

Query Match 95.0%; Score 19; DB 2; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 |||||  
 DB 77 VAEF 80

RESULT 70  
 S10655  
 hypothetical protein X - Pyrococcus woesei (fragment)  
 C/Species: Pyrococcus woesei  
 C/Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 18-Jun-1993  
 C/Accession: S10655

R/Zwickl, P.; Fabry, S.; Bogedain, C.; Haas, A.; Hensel, R.  
 J. Bacteriol. 172, 4329-4338, 1990  
 A/Title: Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeobact  
 scherichia coli.  
 A/Reference number: S10650; MUID:90330536; PMID:2165475  
 A/Accession: S10655  
 A/Molecule type: DNA  
 A/Residues: 1-148 <ZWI>  
 A/Note: the authors translated the codon GGT for residue 54 as Glu

Query Match 95.0%; Score 19; DB 2; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 |||||  
 DB 115 VAEF 118

## RESULT 71

AC1137  
 Ribose 5-phosphate isomerase homolog lmo0498 [imported] - Listeria monocytogenes (strai  
 C/Species: Listeria monocytogenes  
 C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
 C/Accession: AC1137  
 R/Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, F  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A/Authors: Kreft, J.; Kuhn, M.; Kunet, F.; Kurapkat, G.; Maduero, E.; Maitournam, A.; N  
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland  
 A/Title: Comparative genomics of Listeria species.  
 A/Reference number: AB1077; MUID:21537279; PMID:11679669  
 A/Accession: AC1137

A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-149 <GLA>

A/Cross-references: GB:NC\_003210; PIDN:CAC98577.1; PID:g16409874; GSPDB:GN00177  
 A/Experimental source: strain EGB-e  
 C/Genetics:  
 A/Gene: lmo0498  
 C/Superfamily: galactoside O-acetyltransferase

Query Match 95.0%; Score 19; DB 2; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 |||||  
 DB 124 VAEF 127

## RESULT 72

AB2857  
 pseudouridine [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C/Accession: AB2857  
 R/Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCiel  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A/Reference number: AB2577; MUID:21608550; PMID:11743193  
 A/Accession: AB2857

A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-150 <KUR>

A/Cross-references: GB:AE008688; PIDN:AAL43272.1; PID:g17740759; GSPDB:GN00186  
 A/Experimental source: strain C58 (Dupont)  
 C/Genetics:

2 VAEF 5 3v

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: May 24, 2004, 17:33:28 ; Search time 8.57143 Seconds  
(without alignments)  
30.374 Million cell updates/sec

Title: US-09-594-978A-3

Perfect score: 20

Sequence: 1 XVAEF 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	19	95.0	19	1	NS2_MYCTU	P81136	mycobacteri
2	19	95.0	59	1	SECE_TREPA	O83263	treponema p
3	19	95.0	59	1	YA90_ARCFU	O29175	archaeoglob
4	19	95.0	63	1	Y451_HAETN	P43998	haemophilus
5	19	95.0	88	1	R151_HAETN	P44399	haemophilus
6	19	95.0	93	1	ALB2_WETKA	O84266	methanopyru
7	19	95.0	108	1	YF24_METUA	O38919	methanococc
8	19	95.0	114	1	KEDA_ACTSL	P41249	actinomycet
9	19	95.0	115	1	INS_VERMO	Q9w7r2	verasper mo
10	19	95.0	116	1	INS_LOEPI	P01341	lophius pis
11	19	95.0	119	1	U275_DROME	Q9v8f3	drosofila
12	19	95.0	120	1	GTR2_PIG	O62786	sus scrofa
13	19	95.0	122	1	R18E_PYRAE	Q82y42	pyrobaculum
14	19	95.0	124	1	Y670_PASMU	Q9cm50	pasteurella
15	19	95.0	127	1	ILBP_MOUSE	P51162	mus musculus
16	19	95.0	130	1	SPEH_THEMEA	Q9wzc3	thermotoga
17	19	95.0	131	1	Y194_PYRAB	Q9v280	pyrococcus
18	19	95.0	133	1	R15B_HALNI	Q9hms5	halobacteri
19	19	95.0	134	1	YBEC_ECOLI	O89999	escherichia
20	19	95.0	139	1	R15B_WETTH	O27443	methanobact
21	19	95.0	140	1	Y586_ARCFU	O29669	archaeoglob
22	19	95.0	143	1	R15B_ARCFU	O28152	archaeoglob
23	19	95.0	145	1	RL13_HALMA	P29198	haloarcula
24	19	95.0	148	1	CLM4_MOUSE	Q9jtm3	mus musculus
25	19	95.0	148	1	YORX_PYRMO	Q20298	pyrococcus
26	19	95.0	150	1	R15B_PYRAE	Q242e3	pyrobaculum
27	19	95.0	152	1	SODC_PRIGL	P14181	prionace gl
28	19	95.0	154	1	R15B_SULSO	Q980B5	sulfolobus
29	19	95.0	155	1	PCP_HAETN	P10325	haemophilus
30	19	95.0	157	1	RAP_TAROF	O49065	taraxacum o
31	19	95.0	157	1	R15B_SULTO	Q975m5	sulfolobus
32	19	95.0	163	1	SGS5_DROME	P07701	drosofila
33	19	95.0	166	1	UTR5_YEAST	P32630	saccharomyc

34	19	95.0	169	1	SSB_RHOBA	P59932	rhodopirell
35	19	95.0	172	1	BFL1_MOUSE	Q07440	mus musculus
36	19	95.0	175	1	BFL1_HUMAN	Q16548	homo sapien
37	19	95.0	176	1	FR12_RANCA	P07798	rana catesb
38	19	95.0	177	1	RL10_THETN	Q8r7u4	thermoanaer
39	19	95.0	178	1	DSB8_PASMU	Q916b3	pasteurella
40	19	95.0	178	1	ESMS_DROME	P13096	drosofila
41	19	95.0	178	1	YB18_ARCFU	O29147	archaeoglob
42	19	95.0	179	1	AFT_HAEDU	P18902	bos taurus
43	19	95.0	183	1	RETB_BOVIN	P35907	wolbachia s
44	19	95.0	186	1	DNAA_WOLSP	P05417	paracoccu
45	19	95.0	190	1	UCRI_PARDE	Q57679	methanococ
46	19	95.0	193	1	NTPA_METJA	Q8k8d9	bacillus ha
47	19	95.0	194	1	HAM2_BACHD	P05823	escherichia
48	19	95.0	196	1	TNR0_ECOLI	P77170	escherichia
49	19	95.0	196	1	PINQ_ECOLI	P77574	escherichia
50	19	95.0	196	1	PINR_ECOLI	P10940	escherichia
51	19	95.0	196	1	UHPA_ECOLI	P27667	salmonella
52	19	95.0	196	1	UHPA_SALTY	Q9wve0	mus musculu
53	19	95.0	198	1	AICD_MOUSE	Q33777	sulfolobus
54	19	95.0	199	1	HISS_SULSO	Q28369	equus cabal
55	19	95.0	201	1	RETB_HORSE	P02753	homo sapien
56	19	95.0	201	1	RETB_HUMAN	P27485	sus scrofa
57	19	95.0	201	1	RETB_PIG	P06912	oryctolagus
58	19	95.0	201	1	RETB_RABIT	P27862	escherichia
59	19	95.0	204	1	YIGZ_ECOLI	Q13257	homo sapien
60	19	95.0	205	1	MD21_HUMAN	Q921b5	mus musculu
61	19	95.0	205	1	MD21_MOUSE	P37033	legionella
62	19	95.0	208	1	YAC1_LEGFN	P71119	corynebacte
63	19	95.0	215	1	HMUO_CORDI	Q66120	cucumber mo
64	19	95.0	218	1	COAT_CMVIX	O10274	orgyia pseu
65	19	95.0	220	1	PTPI_NPVOP	O29758	archaeoglob
66	19	95.0	223	1	ECRI_ARCFU	P41914	tetrahymena
67	19	95.0	223	1	RAN_TETPY	O87880	thermus the
68	19	95.0	223	1	VATD_THETH	P41915	tetrahymena
69	19	95.0	225	1	RAN_TETTH	P24080	orgyia pseu
70	19	95.0	228	1	Y132_NPVOP	Q88dl5	pseudomonas
71	19	95.0	230	1	NADD_PSEPK	P51308	porphyra pu
72	19	95.0	230	1	RR3_FORPU	Q8f8b3	corynebacte
73	19	95.0	230	1	UBIE_COREF	Q8nt39	corynebacte
74	19	95.0	230	1	YBIE_COREF	P24195	escherichia
75	19	95.0	234	1	YGIB_ECOLI	Q7m9x5	wolinnella s
76	19	95.0	236	1	PUR7_WOLSU	Q50740	mycobacteri
77	19	95.0	236	1	YP58_MYCTU	Q8ctk9	staphylococ
78	19	95.0	238	1	Y437_STAEP	Q8nkr2	staphylococ
79	19	95.0	238	1	Y669_STAAM	O8dsj4	streptococc
80	19	95.0	238	1	YH89_STRMU	Q9j5m7	neisseria m
81	19	95.0	239	1	RLPA_NEIMA	Q91k40	neisseria m
82	19	95.0	239	1	RLPA_NEIME	Q53305	mycobacteri
83	19	95.0	239	1	VG88_BPMLS	Q98j12	arabidopsis
84	19	95.0	240	1	ATP7_ARATH	P58551	pasteurella
85	19	95.0	240	1	KDKA_PASPI	Q93t15	prochloroc
86	19	95.0	241	1	PCYA_PROMP	O66957	aquifex aeo
87	19	95.0	242	1	Y755_AQUAE	Q8dl85	synchococc
88	19	95.0	244	1	COMB_SYNEL	Q07300	orghum bic
89	19	95.0	247	1	ATPD_SORBI	Q87rj6	vibrio para
90	19	95.0	247	1	CYSZ_VIBPA	Q9nhw7	aedes aegyp
91	19	95.0	249	1	AQP_AEDAE	P21653	nicotiana t
92	19	95.0	250	1	TIP1_TOBAC	P24422	nicotiana t
93	19	95.0	250	1	TIP2_TOBAC	P33560	antirrhinum
94	19	95.0	250	1	TIP2_ANTWA	P37769	escherichia
95	19	95.0	253	1	KUUD_ECOLI	Q05528	erwinia chr
96	19	95.0	253	1	KUUD_ERWCH	Q28439	archaeoglob
97	19	95.0	255	1	PPI_BRANA	O15400	homo sapien
98	19	95.0	255	1	TRMB_PASMU	O70439	mus musculu
99	19	95.0	257	1	PSD8_HUMAN	Q9c1c2	pasteurella
100	19	95.0	257	1	PSD8_MOUSE	P48556	homo sapien
101	19	95.0	258	1	NADC_ARCFU	Q9cx56	mus musculu
102	19	95.0	261	1	STX7_HUMAN	O28439	archaeoglob
103	19	95.0	261	1	STX7_MOUSE	O15400	homo sapien
104	19	95.0	261	1	STX7_RAT	O70257	rattus norv
105	19	95.0	266	1	TRPA_SYNEL	Q8dl99	synchococc
106	19	95.0	267	1	APHC_MOUSE	Q9d099	mus musculu

107	19	95.0	1	DKGB_YERPE	Q8zh36	versinia pe	180	19	95.0	325	1	UCP5_MOUSE	Q922b2	mus musculus	
108	19	95.0	267	1	GNPI_CAEEL	Q9xvj2	caenorhabdi	181	19	95.0	326	1	PE46_ARATH	O81772	arabidopsis
109	19	95.0	267	1	MM07_RAT	P50280	rattus norv	182	19	95.0	326	1	YP87_CAEEL	Q94444	caenorhabdi
110	19	95.0	269	1	AQPI_HUMAN	P29972	homo sapien	183	19	95.0	328	1	AIPL_RAT	Q91l99	rattus norv
111	19	95.0	269	1	AQPI_MOUSE	Q02013	mus musculus	184	19	95.0	328	1	Y778_AQUAE	O66974	aquifex aeo
112	19	95.0	269	1	AQPI_RAT	P29975	rattus norv	185	19	95.0	329	1	YFV6_SCHPO	Q9p7q8	schizosacch
113	19	95.0	269	1	RM09_YEAST	P31334	saccharomyc	186	19	95.0	330	1	PRC2_PORGI	P33437	porphyromon
114	19	95.0	269	1	SRPB_MOUSE	P47758	mus musculus	187	19	95.0	330	1	Y840_METUA	Q58250	methanococc
115	19	95.0	271	1	YS13_STAAU	O05337	staphylococ	188	19	95.0	334	1	PRC1_PORGI	P51444	mechanococc
116	19	95.0	271	1	AQPI_BOVIN	P47865	bos taurus	189	19	95.0	334	1	YJH3_RHIME	P59916	porphyromon
117	19	95.0	271	1	SRPE_HUMAN	O9v5m8	homo sapien	190	19	95.0	334	1	YJH3_RHIME	Q92v44	rhizobium m
118	19	95.0	272	1	AQPI_SHEEP	P56401	ovis aries	191	19	95.0	337	1	OPCM_CHICK	Q98892	gallus gall
119	19	95.0	272	1	YF63_STAAM	Q99tt3	staphylococ	192	19	95.0	338	1	Y4XL_RHIN	P55704	rhizobium s
120	19	95.0	274	1	AMYC_THETU	P37729	thermoanaer	193	19	95.0	339	1	HRDC_STROO	P18184	streptomyce
121	19	95.0	274	1	Y087_THEMEA	Q9wxu2	thermotoga	194	19	95.0	339	1	IGBI_HUMAN	P78318	homo sapien
122	19	95.0	277	1	PHEA_LACLC	Q9ceu2	lactococcus	195	19	95.0	340	1	HUR_STRAU	Q00923	streptomyce
123	19	95.0	279	1	PHEA_LACLC	P43909	lactococcus	196	19	95.0	340	1	IGBI_MOUSE	Q61249	mus musculus
124	19	95.0	282	1	TRUB_BORBU	O51743	borrelia bu	197	19	95.0	340	1	IGBI_RAT	O08836	rattus norv
125	19	95.0	284	1	Y3GA_SHEON	Q8ae2	shewanella	198	19	95.0	343	1	MO2M_ARATH	Q9m0m4	arabidopsis
126	19	95.0	284	1	YN13_MYCTU	P71899	mycobacteri	199	19	95.0	343	1	MO2M_ARATH	Q9fgk3	arabidopsis
127	19	95.0	285	1	YB46_HAEIN	P45071	haemophilus	200	19	95.0	343	1	XT30_ARATH	Q38908	arabidopsis
128	19	95.0	287	1	PDXV_ECOLI	P77150	escherichia	201	19	95.0	344	1	CD2_MOUSE	P08920	mus musculus
129	19	95.0	287	1	SYQ_DICDI	P14325	dictyosteli	202	19	95.0	344	1	CD2_MOUSE	P08921	rattus norv
130	19	95.0	288	1	Y169_PASMU	Q9p85	pasteurella	203	19	95.0	344	1	STSY_RAUSE	P53224	rauvolfia s
131	19	95.0	291	1	AMPN_ARCFU	O28438	archaeoglob	204	19	95.0	345	1	TRPD_SULSO	P50384	sulfolobus
132	19	95.0	292	1	AMPN_METJA	Q59023	methanococc	205	19	95.0	347	1	FLIG_AGR75	Q44458	agrobacteri
133	19	95.0	294	1	SSBI_YEAST	Q58725	methanococc	206	19	95.0	352	1	CYCR_ROSDE	P26278	roseobacter
134	19	95.0	294	1	PRVA_VIBCH	P10080	saccharomyc	207	19	95.0	352	1	CYCR_CATRO	P18417	catharanthu
135	19	95.0	295	1	RPBA_XANCP	Q9kv64	vibrio chol	208	19	95.0	354	1	DDI1_HUMAN	P29894	homo sapien
136	19	95.0	295	1	YDH2_XANAU	P52256	xanthomonas	209	19	95.0	362	1	MURG_SHEON	P59894	homo sapien
137	19	95.0	295	1	YD67_METJA	P22644	xanthobacte	210	19	95.0	364	1	DDL_CHLTE	Q8ocr8	chlorobium
138	19	95.0	297	1	T2RI_HUMAN	Q58762	methanococc	211	19	95.0	364	1	DDL_CHLTE	Q8ocr8	chlorobium
139	19	95.0	299	1	NARE_CHICK	Q9nyw7	homo sapien	212	19	95.0	364	1	MRP_AQUAE	O66946	aquifex aeo
140	19	95.0	300	1	CYSD_XYLFA	Q92080	gallus gall	213	19	95.0	365	1	MURG_SHEVI	O9f1n0	shewanella
141	19	95.0	304	1	CYSD_XYLFA	Q9p079	xyliella fas	214	19	95.0	366	1	CYCR_RHOGE	P51758	rhodocyclu
142	19	95.0	304	1	CYSD_XYLFA	Q87d98	xyliella fas	215	19	95.0	369	1	RPSD_STRPN	O03388	streptococc
143	19	95.0	305	1	B2BA_CAEEL	P34604	caenorhabdi	216	19	95.0	378	1	KLAB_ECOLI	Q52328	escherichia
144	19	95.0	308	1	CP93_ENTHI	O06964	entamoeba h	217	19	95.0	379	1	TGT_VIBCH	Q9kty9	vibrio chol
145	19	95.0	308	1	KSGL_BIFLO	Q866i3	b dimethyla	218	19	95.0	380	1	METB_HELPY	Q92mw7	helicobacte
146	19	95.0	309	1	DNUL_MYCPN	Q50312	mycoplasma	219	19	95.0	380	1	METB_HELPY	P55069	helicobacte
147	19	95.0	309	1	NTCB_SYNY3	P74422	synechocyst	220	19	95.0	381	1	CD84_SCHPO	O09184	schizosacch
148	19	95.0	310	1	MIAA_CLOPE	Q8x185	clostridium	221	19	95.0	383	1	CYCR_CHRVI	Q82947	chromatium
149	19	95.0	310	1	PP12_TOBAC	O04857	nicotiana t	222	19	95.0	384	1	AIPL_HUMAN	Q95mn9	pan paniscu
150	19	95.0	310	1	SIR5_HUMAN	Q9nxa8	homo sapien	223	19	95.0	384	1	AIPL_PANFA	Q20701	caenorhabdi
151	19	95.0	310	1	SIR5_MOUSE	Q8k2c6	mus musculus	224	19	95.0	385	1	GBA5_CAEEL	P43726	haemophilus
152	19	95.0	310	1	Y942_SULSO	Q97zh1	sulfolobus	225	19	95.0	385	1	PGK_HAEIN	P10312	bacterioph
153	19	95.0	311	1	DAPA_STRPN	Q97r25	streptococc	226	19	95.0	387	1	VPD_BPP2	P10312	bacterioph
154	19	95.0	311	1	DAPA_STRPN	Q8dpz9	streptococc	227	19	95.0	388	1	METB_MYCLE	P46807	mycobacteri
155	19	95.0	311	1	MIAA_LACPL	Q88wp5	lactobacill	228	19	95.0	388	1	METB_MYCTU	O53427	mycobacteri
156	19	95.0	311	1	O5L1_HUMAN	Q8ng12	homo sapien	229	19	95.0	389	1	RFLM_CAEEL	O44568	caenorhabdi
157	19	95.0	311	1	O5L2_HUMAN	Q8ng10	homo sapien	230	19	95.0	389	1	VPD_BF186	P21679	bacterioph
158	19	95.0	312	1	O6N1_HUMAN	Q8gy5	homo sapien	231	19	95.0	390	1	Y181_ARCFU	O30056	archaeoglob
159	19	95.0	312	1	PP12_ARATH	P48482	arabidopsis	232	19	95.0	392	1	TEBB_STYMY	P29548	stylonchla
160	19	95.0	312	1	PP15_ARATH	P48485	arabidopsis	233	19	95.0	392	1	YBUJ_HAEIN	P44083	haemophilus
161	19	95.0	313	1	DRPE_CRAPL	P22242	craterostig	234	19	95.0	394	1	DXR_SYNY3	Q55663	synechocyst
162	19	95.0	313	1	PEPM_STRHY	P29247	streptomyce	235	19	95.0	395	1	CGA2_CHICK	P43449	gallus gall
163	19	95.0	314	1	IUNH_CRIFA	Q27546	crithidia f	236	19	95.0	397	1	Y923_THEMA	P56727	thermotoga
164	19	95.0	314	1	PP1_PHAVU	P48490	phaseolus v	237	19	95.0	398	1	FNGB_BUCAL	Q8x916	buchnera ap
165	19	95.0	315	1	O3A2_HUMAN	P47893	homo sapien	238	19	95.0	398	1	FNGB_BUCAL	P57442	buchnera ap
166	19	95.0	315	1	O3A2_PANTR	Q9tu97	pan troglod	239	19	95.0	400	1	ER19_HUMAN	P53602	homo sapien
167	19	95.0	315	1	O3A3_HUMAN	P47888	homo sapien	240	19	95.0	401	1	ER19_MOUSE	Q99j15	mus musculus
168	19	95.0	315	1	O3A3_PANTR	Q9tuao	pan troglod	241	19	95.0	401	1	ER19_MOUSE	Q62967	rattus norv
169	19	95.0	316	1	APEI_MOUSE	P28352	mus musculus	242	19	95.0	402	1	CL13_MYCTU	P77902	mycobacteri
170	19	95.0	316	1	APEI_RAT	P43138	rattus norv	243	19	95.0	402	1	CL13_XYLFA	Q99GC5	xyliella fas
171	19	95.0	316	1	PP1_BRAOL	P48487	brassica ol	244	19	95.0	402	1	CL13_XYLFA	Q87ax5	xyliella fas
172	19	95.0	317	1	APEI_HUMAN	P27695	homo sapien	245	19	95.0	406	1	CGA2_BOVIN	P30274	bos taurus
173	19	95.0	317	1	O5TF_RAT	P08721	rattus norv	246	19	95.0	406	1	CGA2_BOVIN	P48775	homo sapien
174	19	95.0	317	1	PP11_TOBAC	O04856	nicotiana t	247	19	95.0	407	1	T250_HUMAN	O60858	homo sapien
175	19	95.0	318	1	PP11_ARATH	P30366	arabidopsis	248	19	95.0	409	1	Y4DM_RHIN	P55412	rhizobium s
176	19	95.0	322	1	PP16_ARATH	P48486	arabidopsis	249	19	95.0	410	1	TDPI_HUMAN	Q14186	homo sapien
177	19	95.0	324	1	ADD_RHIME	Q92t48	rhizobium m	250	19	95.0	410	1	TDPI_MOUSE	Q08639	mus musculus
178	19	95.0	324	1	PP18_ARATH	O82734	arabidopsis	251	19	95.0	411	1	PEPT_YERPE	Q8zfr0	yersinia pe
179	19	95.0	325	1	UCP5_HUMAN	O95258	homo sapien	252	19	95.0	412	1	CCA_ECOLI	P06961	escherichia



399	19	95.0	580	1	RS1_CHLPN	Q928m3 chlamydia p	472	19	95.0	649	1	INVA_PHAU	P29001 phaseolus a
400	19	95.0	587	1	COBE_ONCMY	Q90x85 oncorhynch	473	19	95.0	651	1	INVA_PHAU	Q24509 phaseolus v
401	19	95.0	588	1	HASI_XENLA	P33563 xenopus lae	474	19	95.0	652	1	C3BB_BACTU	Q06117 bacillus th
402	19	95.0	589	1	GBP1_MOUSE	Q01514 mus musculu	475	19	95.0	660	1	FDL_DROME	Q08w13 drosophila
403	19	95.0	589	1	GBP2_MOUSE	Q02066 mus musculu	476	19	95.0	660	1	TP6B_METJA	Q58434 methanococc
404	19	95.0	589	1	GBP2_RAT	Q63663 rattus norv	477	19	95.0	676	1	NICA_ARATH	Q8gum5 arabidopsis
405	19	95.0	591	1	GBP2_HUMAN	P32456 homo sapien	478	19	95.0	677	1	NCPR_MOUSE	P37040 mus musculu
406	19	95.0	592	1	ALAM_YEAST	P22893 saccharomyc	479	19	95.0	677	1	NCPR_RAT	P00388 rattus norv
407	19	95.0	602	1	PEV5_HUMAN	P50545 homo sapien	480	19	95.0	680	1	DCP_SALTY	P27236 salmonella
408	19	95.0	602	1	TP6B_ARCFU	Q29605 archaeoglob	481	19	95.0	685	1	CSD_MYCPA	Q9k1i6 mycobacteri
409	19	95.0	604	1	DNAB_BACME	P05646 bacillus me	482	19	95.0	690	1	TRP1_SALSA	P80426 salmo salar
410	19	95.0	605	1	APM2_YEAST	P38700 saccharomyc	483	19	95.0	691	1	TRP2_SALSA	P80429 salmo salar
411	19	95.0	607	1	BKL3_YEAST	Q8bh14 mus musculu	484	19	95.0	692	1	HEXA_ANOGA	Q17020 anopheles g
412	19	95.0	607	1	DNAB_FUSIN	Q8rh05 fusobacteri	485	19	95.0	693	1	REGC_PASMU	Q9cm44 pasteurella
413	19	95.0	607	1	DNAB_LACIA	P42368 lactococcus	486	19	95.0	695	1	TRPE_RABIT	Q17020 anopheles g
414	19	95.0	607	1	DNAB_STRP3	Q8k624 streptococc	487	19	95.0	696	1	TRPE_FIG	P19134 oryctolagus
415	19	95.0	607	1	DNAB_STRP3	P58829 streptococc	488	19	95.0	698	1	TRPE_HUMAN	P09571 sus scrofa
416	19	95.0	607	1	DNAB_STRP3	Q8wt3 streptococc	489	19	95.0	704	1	ICA_FIG	Q2787 homo sapien
417	19	95.0	607	1	DNAB_STRP6	Q8wt3 streptococc	490	19	95.0	704	1	TRPE_BOVIN	Q29545 sus scrofa
418	19	95.0	608	1	BKL3_HUMAN	Q8ab22 homo sapien	491	19	95.0	705	1	BPG_BACTN	Q29443 bos taurus
419	19	95.0	609	1	DNAB_ENTFA	Q83577 enterococcu	492	19	95.0	705	1	PAL2_LITER	Q8a474 bacteroides
420	19	95.0	610	1	DNAB_BACSH	Q59268 bacillus sp	493	19	95.0	706	1	DREB_MOUSE	Q49836 lithospermu
421	19	95.0	610	1	DNAB_STRAH	P45554 staphylococ	494	19	95.0	706	1	EPG_LEPIN	Q9qxs6 mus musculu
422	19	95.0	610	1	DNAB_STRAH	Q81182 staphylococ	495	19	95.0	707	1	DREB_HORSE	P27425 equus cabal
423	19	95.0	611	1	DNAB_BACAA	Q81182 staphylococ	496	19	95.0	707	1	DREB_RAT	Q07266 rattus norv
424	19	95.0	611	1	DNAB_BACOR	Q81182 staphylococ	497	19	95.0	707	1	PAL1_IPOBA	P14166 ipomoea bat
425	19	95.0	612	1	DNAB_LISIN	Q81182 staphylococ	498	19	95.0	708	1	PAL1_IPOBA	Q23865 daucus caro
426	19	95.0	612	1	DNAB_LISIN	Q81182 staphylococ	499	19	95.0	708	1	TALA_IPOBA	Q42858 ipomoea bat
427	19	95.0	612	1	ILVD_XANAC	Q8p554 xanthomonas	500	19	95.0	708	1	TALA_SV40	P03070 simian viru
428	19	95.0	612	1	YCB1_MYCTU	Q8p554 xanthomonas	501	19	95.0	710	1	PAL1_LITER	Q49836 lithospermu
429	19	95.0	612	1	YCB1_MYCTU	Q8p554 xanthomonas	502	19	95.0	712	1	PAL2_TOBAC	Q49836 lithospermu
430	19	95.0	614	1	DNAB_LACSN	Q8k166 lactobacill	503	19	95.0	712	1	PAL3_TOBAC	P35513 nicotiana t
431	19	95.0	614	1	SPKE_SYNT3	P73515 synchocyst	504	19	95.0	714	1	ZW10_DROGR	Q44219 drosophila
432	19	95.0	621	1	TP6B_METWA	Q8p554 xanthomonas	505	19	95.0	715	1	FAOB_PSEFR	P28793 p fatty oxi
433	19	95.0	621	1	TP6B_METWA	Q8p554 xanthomonas	506	19	95.0	716	1	SYGB_MYCTU	Q8p554 xanthomonas
434	19	95.0	622	1	C2AC_BACTU	Q8p554 xanthomonas	507	19	95.0	721	1	GLGX_MYCTU	Q8p554 xanthomonas
435	19	95.0	624	1	COOS_METJA	Q58138 methanococc	508	19	95.0	722	1	SYGB_SYNT3	Q8p554 xanthomonas
436	19	95.0	624	1	P212_HUMAN	Q9nm26 homo sapien	509	19	95.0	725	1	PAL1_ARATH	P35513 nicotiana t
437	19	95.0	625	1	TPB2_HAEIN	P44971 haemophilus	510	19	95.0	725	1	VRZA_BPT4	P03690 bacterioph
438	19	95.0	627	1	CACP_CANTR	P52826 columba liv	511	19	95.0	729	1	YFPA_ECOLI	P23842 escherichia
439	19	95.0	627	1	CACP_CANTR	Q00614 candida tro	512	19	95.0	759	1	PURL_CHLTF	Q8k17 chlorobium
440	19	95.0	631	1	NTF1_VACCA	P52826 columba liv	513	19	95.0	763	1	P8RA_WOLVS	P31075 wolfinella s
441	19	95.0	631	1	NTF1_VACCA	Q57214 vaccinia vi	514	19	95.0	764	1	ACOC_CUCMC	Q42669 cucumis mel
442	19	95.0	631	1	NTF1_VACCV	P20637 vaccinia vi	515	19	95.0	781	1	DPOL_ARCFU	Q29753 archaeoglob
443	19	95.0	631	1	NTF1_VACCV	P05807 vaccinia vi	516	19	95.0	781	1	OXAA_CHLMU	Q29753 archaeoglob
444	19	95.0	632	1	NTF1_MYXVL	P33066 variola vir	517	19	95.0	787	1	OXAA_CHLTR	Q84253 chlamydia m
445	19	95.0	632	1	NTF1_SFVKA	Q9q814 myxoma viru	518	19	95.0	787	1	YD61_SCHPO	Q10313 schizosacch
446	19	95.0	634	1	DNAB_HAEIN	Q9q822 Shope fibro	519	19	95.0	804	1	TP6B_HALN1	Q9hr31 halobacteri
447	19	95.0	634	1	NTF1_MCVI	P43736 haemophilus	520	19	95.0	807	1	ANCA_MOUSE	Q91w96 mus musculu
448	19	95.0	635	1	DNAB_VIBCH	Q98267 molluscum c	521	19	95.0	808	1	ANCA_HUMAN	Q91u55 homo sapien
449	19	95.0	635	1	DXS_ANASP	Q34241 vibrio chol	522	19	95.0	808	1	GCS1_SCHPO	Q14255 schizosacch
450	19	95.0	636	1	DNAB_VIBVU	Q8y280 anabaena sp	523	19	95.0	818	1	MUTS_THETH	Q56239 thermus the
451	19	95.0	636	1	DNAB_VIBVU	Q8df66 vibrio vuln	524	19	95.0	821	1	CAN3_HUMAN	P20807 homo sapien
452	19	95.0	636	1	MYB_MOUSE	Q7m885 vibrio vuln	525	19	95.0	821	1	SYFB_SYNEL	Q8dl37 synchocyst
453	19	95.0	637	1	DNAB_BRUME	P06976 mus musculu	526	19	95.0	835	1	UL52_HSV8A	P14346 herpesviru
454	19	95.0	637	1	DNAB_BRUME	Q8y876 brucella me	527	19	95.0	842	1	HIS2_PICPA	P45353 pichia past
455	19	95.0	637	1	DNAB_BRUSU	Q05981 brucella ov	528	19	95.0	852	1	CLPB_CORGL	P53532 corynebacte
456	19	95.0	637	1	DNAB_ECOLI	Q8fx22 brucella su	529	19	95.0	858	1	CHS1_RHOL	P30594 rhizopus ol
457	19	95.0	637	1	DNAB_PSEAE	P04475 escherichia	530	19	95.0	861	1	PO58_CAEEL	P34552 caenorhabdi
458	19	95.0	637	1	DNAB_SALTI	Q9hv43 pseudomonas	531	19	95.0	862	1	MUTS_BORBU	O17373 borrelia bu
459	19	95.0	637	1	DNAB_SALTY	Q56073 salmonella	532	19	95.0	876	1	VP41_CASEL	O19954 caenorhabdi
460	19	95.0	637	1	DNAB_VIBPA	Q87x33 vibrio para	533	19	95.0	877	1	DFOL_LACLA	Q9cd81 lactococcus
461	19	95.0	637	1	DNAB_VIBPA	Q91721 vibrio prot	534	19	95.0	877	1	DFOL_LACLC	Q32801 lactococcus
462	19	95.0	638	1	DNAB_EUCEP	P59865 buchnera ap	535	19	95.0	877	1	DFOL_STRPN	P59199 streptococc
463	19	95.0	638	1	DNAB_RHILE	Q33528 rhizobium l	536	19	95.0	881	1	HELI_HSV2H	P28277 herpes simp
464	19	95.0	640	1	DNAB_SHIFL	Q83mh5 shigella fl	537	19	95.0	882	1	SYA_SYNEL	Q8dh56 synchocyst
465	19	95.0	640	1	DNAB_VIBHA	Q87384 vibrio harv	538	19	95.0	889	1	SVX_HUMAN	Q9un79 homo sapien
466	19	95.0	640	1	MYB_BOVIN	P46200 bos taurus	539	19	95.0	889	1	SEC3_HUMAN	Q9nv70 homo sapien
467	19	95.0	642	1	MYB_HUMAN	Q10242 homo sapien	540	19	95.0	894	1	SEC3_MOUSE	Q8r386 mus musculu
468	19	95.0	642	1	INV5_VICFA	Q43857 vicia faba	541	19	95.0	895	1	PNM1_CANAL	P28877 candida alb
469	19	95.0	644	1	H570_BRUMA	P27541 brugia mala	542	19	95.0	899	1	YMU3_YEAST	Q04500 saccharomyc
470	19	95.0	644	1	REB2_RHARA	P17279 rhizomucor	543	19	95.0	916	1	PNM1_AJECA	Q07421 ajellomyces
471	19	95.0	649	1	DREB_HUMAN	Q16643 homo sapien	544	19	95.0				

545	1	SVL STAEF	916	95.0	19	618	19	95.0	2226	1	POLG_HPAV4	P26581	hepatitis a
546	1	Y853_HUMAN	919	95.0	19	619	19	95.0	2226	1	POLG_HPAV8	P26582	hepatitis a
547	1	DMF1_SCHPO	920	95.0	19	620	19	95.0	2227	1	POLG_HPAVH	P08617	hepatitis a
548	1	PMAL_NEUCR	921	95.0	19	621	19	95.0	2227	1	POLG_HPAVL	P06441	hepatitis a
549	1	SYI_BACSU	921	95.0	19	622	19	95.0	2230	1	POLG_HPAVS	P14553	simian hepa
550	1	DPO1_RICFE	922	95.0	19	623	19	95.0	2314	1	PTFZ_HUMAN	P23471	homo sapien
551	1	YDM6_SCHPO	929	95.0	19	624	19	95.0	2316	1	PTFZ_RAT	P26266	rattus norv
552	1	MSH2_ARATH	937	95.0	19	625	19	95.0	2649	1	P285_HUMAN	Q9byk8	homo sapien
553	1	SYL_METH	937	95.0	19	626	19	95.0	2725	1	FLNC_HUMAN	P14315	homo sapien
554	1	MSH2_WAIZE	942	95.0	19	627	19	95.0	2774	1	MAPA_RAT	P34926	rattus norv
555	1	LYAG_HUMAN	952	95.0	19	628	19	95.0	2805	1	MAPA_HUMAN	P78559	homo sapien
556	1	SYV_VIBCH	953	95.0	19	629	19	95.0	3412	1	POLG_TBEVS	P07720	t genome po
557	1	VIL3_ARATH	966	95.0	19	630	19	95.0	3414	1	POLG_LANVT	P23837	t genome po
558	1	SYL_MYCTU	969	95.0	19	631	19	95.0	3414	1	POLG_TBEVH	P14336	t genome po
559	1	AMPN_HAECO	971	95.0	19	632	19	95.0	3415	1	POLG_TBEVW	Q4158	t genome po
560	1	MTRA_YEAST	972	95.0	19	633	19	95.0	3803	1	TRAI_DROME	Q818u7	drosophila
561	1	UVRA_RHIL0	973	95.0	19	634	19	95.0	3859	1	TRAP_HUMAN	Q9y4as	homo sapien
562	1	POD1_CAUCR	974	95.0	19	635	19	95.0	4447	1	PKSK_BACSU	P40803	bacillus su
563	1	VIL2_ARATH	976	95.0	19	636	19	95.0	5179	1	MUC2_HUMAN	Q02817	homo sapien
564	1	POLG_LIV	980	95.0	19	637	19	95.0	5217	1	MLL2_HUMAN	Q1886	cochliobolu
565	1	UVRA_XANCP	987	95.0	19	638	19	95.0	5262	1	MDN1_HUMAN	Q4686	homo sapien
566	1	UVRA_MICLU	988	95.0	19	639	19	95.0	5596	1	CALM_TETH	Q05055	tetrahymena
567	1	CBF_MOUSE	992	95.0	19	640	19	95.0	12	1	YOZD_BACSU	Q31863	bacillus su
568	1	CBF_HUMAN	998	95.0	19	641	19	95.0	58	1	EX7S_STR33	Q8e6m0	streptococc
569	1	DPOG_PICPA	1012	95.0	19	642	19	95.0	71	1	EX7S_STR33	Q8vm65	streptococc
570	1	MAPA_MOUSE	1021	95.0	19	643	19	95.0	79	1	EX7S_LACLA	Q8ch83	lactococcu
571	1	MAPA_MOUSE	1021	95.0	19	644	19	95.0	81	1	YMW4_ECOLI	P76036	escherichia
572	1	PDRI_YEAST	1063	95.0	19	645	19	95.0	88	1	RS15_PHOIU	P41120	photorhabdi
573	1	SMCS_SCHPO	1065	95.0	19	646	19	95.0	92	1	YFCL_ECOLI	P76496	escherichia
574	1	RNC_CAEEL	1086	95.0	19	647	19	95.0	94	1	SELK_HUMAN	Q9y6d0	homo sapien
575	1	AT9B_HUMAN	1095	95.0	19	648	19	95.0	94	1	SELK_MOUSE	Q91j1l	mus muscucu
576	1	AT9B_MOUSE	1095	95.0	19	649	19	95.0	94	1	SELK_MOUSE	P59798	rattus norv
577	1	CYT4_NEUCR	1117	95.0	19	650	19	95.0	94	1	SELK_MOUSE	Q9k5n8	bacillus ha
578	1	V120_HSV11	1123	95.0	19	651	19	95.0	98	1	YCI1_ECOLI	P31070	escherichia
579	1	MAP4_MOUSE	1125	95.0	19	652	19	95.0	101	1	CYAY_HABIN	P71358	haemophilus
580	1	FLNC_MOUSE	1144	95.0	19	653	19	95.0	108	1	PRI8_NITEU	Q82xq7	nitrosomona
581	1	YHC3_YEAST	1155	95.0	19	654	19	95.0	110	1	VAG1_TOBAC	Q82702	nicotiana t
582	1	IF2P_METJA	1155	95.0	19	655	19	95.0	111	1	VAG2_TOBAC	Q82703	nicotiana t
583	1	CAGA_HELPJ	1167	95.0	19	656	19	95.0	117	1	RNPA_LACLA	Q9cj73	lactococcu
584	1	UB4B_MOUSE	1173	95.0	19	657	19	95.0	133	1	S3AD_BACSU	P49781	bacillus su
585	1	MOG5_CABEL	1200	95.0	19	658	19	95.0	134	1	YK49_STRCO	P6251	streptomyce
586	1	DP3A_HELPJ	1211	95.0	19	659	19	95.0	136	1	RL19_XYLEFA	Q9s3m0	bacillus ps
587	1	TOP2_CRIFA	1211	95.0	19	660	19	95.0	136	1	RL19_XYLEFA	Q9ph36	xyella fas
588	1	UB4B_HUMAN	1302	95.0	19	661	19	95.0	141	1	RL19_XYLEFA	Q87f53	xyella fas
589	1	SPR4_CABEL	1311	95.0	19	662	19	95.0	141	1	RL19_XYLEFA	Q02380	tyrophagus
590	1	CYAB_LEIDO	1331	95.0	19	663	19	95.0	141	1	R18B_METJA	Q57751	methanococc
591	1	DP3A_SVNY3	1355	95.0	19	664	19	95.0	144	1	YK14_VIBVU	Q58398	methanococc
592	1	VCAP_HSVSA	1371	95.0	19	665	19	95.0	145	1	YK14_VIBVU	Q8db14	vibrio vuln
593	1	ALBU_PETMA	1423	95.0	19	666	19	95.0	147	1	CALM_KLULA	O60041	kluyveromyc
594	1	NPH4_HUMAN	1426	95.0	19	667	19	95.0	147	1	CALM_YEAST	P06787	saccharomyc
595	1	DPO3_LISIN	1444	95.0	19	668	19	95.0	147	1	YESE_BACSU	O31511	bacillus su
596	1	DPO3_LISMO	1444	95.0	19	669	19	95.0	147	1	YESE_BACSU	P27163	petunia hyb
597	1	YORI_YEAST	1477	95.0	19	670	19	95.0	148	1	CAL2_PETHY	P49258	drosophila
598	1	YORI_YEAST	1477	95.0	19	671	19	95.0	148	1	CAL2_PETHY	P15094	achlya kleb
599	1	YCF1_YEAST	1515	95.0	19	672	19	95.0	148	1	CALM_ACHKL	Q9hfy6	blastocladia
600	1	MRP3_RAT	1522	95.0	19	673	19	95.0	148	1	CALM_ACHKL	P07181	drosophila
601	1	V133_HUMAN	1524	95.0	19	674	19	95.0	148	1	CALM_BLAEM	P02594	electrophor
602	1	MRP3_HUMAN	1527	95.0	19	675	19	95.0	148	1	CALM_BLAEM	P11118	euglena gra
603	1	LMJ1_CABEL	1535	95.0	19	676	19	95.0	148	1	CALM_ELEEL	P13565	hordeum vul
604	1	LHR_ECOLI	1538	95.0	19	677	19	95.0	148	1	CALM_ELEEL	P02593	homo sapien
605	1	CTPI_MYCLE	1609	95.0	19	678	19	95.0	148	1	CALM_HORVU	P27161	lycopersico
606	1	GCC2_MOUSE	1679	95.0	19	679	19	95.0	148	1	CALM_HORVU	P41040	zea mays (m
607	1	YFAO_ANASP	1906	95.0	19	680	19	95.0	148	1	CALM_WAIZE	P02596	metridium s
608	1	MYH9_RAT	1961	95.0	19	681	19	95.0	148	1	CALM_WAIZE	P29612	oryza sativ
609	1	DOC3_MOUSE	2027	95.0	19	682	19	95.0	148	1	CALM_WAIZE	P07463	paramescium
610	1	DOC3_HUMAN	2030	95.0	19	683	19	95.0	148	1	CALM_ORYSA	P02595	patinopecte
611	1	TEGO_HSVJ1	2059	95.0	19	684	19	95.0	148	1	CALM_ORYSA	P27165	phytophthor
612	1	MYOF_HUMAN	2061	95.0	19	685	19	95.0	148	1	CALM_PATSP	P11121	pyridae sp
613	1	RRPL_HRSVA	2165	95.0	19	686	19	95.0	148	1	CALM_PATSP	P13868	solanum tub
614	1	POLN_MANCV	2208	95.0	19	687	19	95.0	148	1	CALM_PATSP	P04353	spinacia ol
615	1	RRPO_TACVC	2210	95.0	19	688	19	95.0	148	1	CALM_PATSP	P21251	stichopus j
616	1	DPOE_YEAST	2222	95.0	19	689	19	95.0	148	1	CALM_PATSP		
617	1	POLG_HPAV2	2226	95.0	19	690	19	95.0	148	1	CALM_PATSP		

691	18	90.0	148	1	CALM_STYLE	P27166 stylonychia	754	18	90.0	229	1	RAD8_THEAC	Q9hid3 thermoplasm
692	18	90.0	148	1	CALM_TETPY	P02598 tetrahymena	785	18	90.0	230	1	ISPD_SYNY3	P74323 synechocyst
693	18	90.0	148	1	CALS_CHICK	P02597 gallus gall	766	18	90.0	231	1	FP_CRIMI	P15697 mesocricetu
694	18	90.0	149	1	CALM_CANAL	P23286 candida alb	767	18	90.0	234	1	SAMP_MESAU	P07629 mesocricetu
695	18	90.0	149	1	CALM_WHEAT	P04464 triticum ae	768	18	90.0	235	1	ICLN_CANFA	P35521 canis famli
696	18	90.0	149	1	MLE3_HUMAN	P06741 homo sapien	769	18	90.0	236	1	PUR7_HELHP	Q7vfq6 helicobacte
697	18	90.0	149	1	MLE3_RABIT	P02603 oryctolagus	770	18	90.0	237	1	PUR7_METAC	Q8ti89 methanosarc
698	18	90.0	149	1	MLE3_RAT	P02601 rattus norv	771	18	90.0	237	1	PUR7_METAC	Q8pyk6 methanosarc
699	18	90.0	150	1	CALM_SCHPO	P05933 schizosacch	772	18	90.0	238	1	RL1_ANASP	Q8vj17 anabaena sp
700	18	90.0	150	1	CALM_SCHPO	P05933 schizosacch	773	18	90.0	238	1	RL1_ANASP	Q8vj17 anabaena sp
701	18	90.0	151	1	X351_METJA	Q57797 methanococc	774	18	90.0	238	1	Y538_CHLTR	Q8pj44 chlamydia t
702	18	90.0	151	1	CP2B_DROME	P02599 dictyosteli	775	18	90.0	240	1	Y825_CHLMU	Q83ca8 coxiella bu
703	18	90.0	151	1	YAG6_RHINE	Q9nlp6 drosophila	776	18	90.0	242	1	MTGA_ECOLI	P46022 escherichia
704	18	90.0	152	1	AAC6_ENTAE	P13486 rhizobium m	777	18	90.0	242	1	MTGA_KLEOX	Q48465 klebsiella
705	18	90.0	155	1	CALF_NABGR	P50858 enterobacte	778	18	90.0	242	1	NPDI_THETN	Q8r9n6 thermosanae
706	18	90.0	155	1	Y463_FUSNN	P53440 naegleria g	779	18	90.0	242	1	PDAXI_NEIMA	Q9rcv9 neisseria m
707	18	90.0	159	1	TPCS_HUMAN	Q8rg55 fusobacteri	780	18	90.0	242	1	TRPC_SCVLA	Q93426 homo sapien
708	18	90.0	159	1	TPCS_MOUSE	P02585 homo sapien	781	18	90.0	244	1	TRCB_HUNAN	Q28148 archaeoglob
709	18	90.0	159	1	TPCS_PIG	P20801 mus musculu	782	18	90.0	244	1	YL32_ARCFU	P75914 escherichia
710	18	90.0	159	1	TPCS_RABIT	P02587 sus scrofa	783	18	90.0	245	1	YCDX_ECOLI	Q26206 methanobact
711	18	90.0	160	1	TPCS_ANGAN	P02586 oryctolagus	784	18	90.0	246	1	AQPM_METTH	Q9c425 methanobact
712	18	90.0	162	1	AROK_LACUA	P81660 anguilla an	785	18	90.0	246	1	AQPM_METTH	Q59335 chlorobium
713	18	90.0	162	1	CALM_CHLRE	Q9ceul lactococcus	786	18	90.0	246	1	HEMA_CHLVI	P44890 haemophilus
714	18	90.0	162	1	TPCS_CHICK	P04352 chlamydomon	787	18	90.0	246	1	MTGA_HAEIN	Q272a1 sulfobus
715	18	90.0	162	1	TPCS_MELGA	P02588 gallus gall	788	18	90.0	248	1	TRPC_SULTO	P55435 rhizobium s
716	18	90.0	162	1	TPCS_RANES	P10246 meleagris g	789	18	90.0	251	1	Y4EL_RHISM	Q87wq1 pseudomonas
717	18	90.0	166	1	Y551_SULSO	P02589 rana escul	790	18	90.0	251	1	TPIS_P3ESM	P95576 pseudomonas
718	18	90.0	166	1	Y228_METJA	Q9uwy6 sulfobus	791	18	90.0	251	1	TPIS_P3ESY	Q927w5 chlamydia p
719	18	90.0	169	1	IPYR_ANGP	Q58045 methanococc	792	18	90.0	252	1	RECO_CHLNP	Q56319 thermotoga
720	18	90.0	169	1	IPYR_ANGP	P80562 anabaena sp	793	18	90.0	252	1	TPPC_THEMA	P53560 bacillus su
721	18	90.0	171	1	NSG2_HUMAN	Q90507 synechocyst	794	18	90.0	253	1	YTBQ_BACSU	Q8jytc0 neisseria m
722	18	90.0	171	1	NSG2_MOUSE	Q97328 homo sapien	795	18	90.0	254	1	AROD_NEIMB	Q97283 saccharomyc
723	18	90.0	171	1	YBP2_ACIAM	P47759 mus musculu	796	18	90.0	254	1	PMW_YEAST	Q9K5j6 bacillus an
724	18	90.0	172	1	IPYR_SYNEL	P32986 acidianus a	797	18	90.0	257	1	RPSE_BACAA	P39161 escherichia
725	18	90.0	173	1	TEPA_MEEME	Q8dht2 synechococc	798	18	90.0	257	1	UXUR_ECOLI	P42608 escherichia
726	18	90.0	173	1	TEPA_MEEME	P50157 abysetoma m	799	18	90.0	258	1	EXUR_ECOLI	Q9x960 erwinia chr
727	18	90.0	174	1	MLRN_DROME	P50422 drosophila	800	18	90.0	259	1	EXUR_ERWCH	Q8y9p15 anabaena sp
728	18	90.0	177	1	IPYR_HALMI	Q9nsf3 halobacteri	801	18	90.0	260	1	RS3_ANASP	Q8a4d6 bacteroides
729	18	90.0	181	1	PYR3_THEAO	P96078 thermus aqu	802	18	90.0	260	1	TRPC_BACTN	P06574 bacillus su
730	18	90.0	184	1	Y803_CHLNP	Q9c7a3 chlamydia p	803	18	90.0	261	1	RPSE_BACSU	P53853 saccharomyc
731	18	90.0	185	1	CBX1_HUMAN	Q277a3 chlamydia t	804	18	90.0	264	1	YNY6_YEAST	Q03768 saccharomyc
732	18	90.0	185	1	DPI_HUMAN	P23197 homo sapien	805	18	90.0	265	1	GIR2_YEAST	P44487 haemophilus
733	18	90.0	185	1	DPI_MOUSE	Q00765 homo sapien	806	18	90.0	266	1	UXUR_HAEIN	Q9nun7 homo sapien
734	18	90.0	185	1	NSG1_HUMAN	Q60870 mus musculu	807	18	90.0	267	1	APHC_HUNAN	Q9uxa6 sulfobus
735	18	90.0	185	1	NSG1_MOUSE	P42857 homo sapien	808	18	90.0	267	1	RL4_SULSO	Q48215 haemophilus
736	18	90.0	185	1	RPFC_ECOLI	Q82092 mus musculu	809	18	90.0	267	1	VG95_HAEIN	Q34411 bacillus su
737	18	90.0	185	1	Y56A_THEMA	P37745 escherichia	810	18	90.0	268	1	HIS9_BACSU	Q28204 archaeoglob
738	18	90.0	193	1	KTHY_THEAC	P58008 thermotoga	811	18	90.0	268	1	THIM_ARCFU	P19867 bacillus st
739	18	90.0	197	1	Y249_METTH	Q9hlx2 thermoplasm	812	18	90.0	269	1	TRPA_BACST	Q28204 archaeoglob
740	18	90.0	199	1	TDX2_BRUMA	Q26351 methanobact	813	18	90.0	270	1	RPNC_SCHPO	P50524 schizosacch
741	18	90.0	200	1	ESTE_VIEMI	Q17172 brugia mala	814	18	90.0	270	1	YC87_METJA	Q58683 methanococc
742	18	90.0	201	1	RETB_MOUSE	Q07792 vibrio mimi	815	18	90.0	272	1	3DHQ_ACICA	Q59087 acinetobact
743	18	90.0	201	1	RETB_MOUSE	Q00724 mus musculu	816	18	90.0	272	1	AQPA_RANES	P50501 rana escul
744	18	90.0	202	1	COAE_CHLFR	P04916 rattus norv	817	18	90.0	275	1	SC65_CANAL	Q14415 candida alb
745	18	90.0	202	1	HIS5_METAC	Q84499 chlamydia t	818	18	90.0	277	1	Y154_ARCFU	Q30083 archaeoglob
746	18	90.0	202	1	HIS5_METMA	Q8ts91 methanosarc	819	18	90.0	278	1	Y128_ARATH	Q92vxs8 arabidopsis
747	18	90.0	202	1	YC55_MCTTU	Q8pvd5 methanosarc	820	18	90.0	279	1	YD15_STAAM	Q8nwr0 staphylococ
748	18	90.0	206	1	CNC2_MOUSE	Q11063 mycobacteri	821	18	90.0	279	1	YE25_STAAM	Q99u62 staphylococ
749	18	90.0	208	1	CNC2_HUMAN	Q9db76 mus musculu	822	18	90.0	280	1	PI27_ARATH	P93004 arabidopsis
750	18	90.0	208	1	RR1_PEA	Q9y3bc homo sapien	823	18	90.0	281	1	GLPF_ECOLI	P31124 escherichia
751	18	90.0	209	1	RR1_PEA	P49208 pisum sativ	824	18	90.0	281	1	GLPF_SHIFL	P31140 shigella fl
752	18	90.0	216	1	TRPF_METKA	P16631 gracilaria	825	18	90.0	282	1	LGT_HELJF	P42787 acirplex ca
753	18	90.0	217	1	DEP1_BIFLO	Q9g534 bifidobacte	826	18	90.0	283	1	LGT_HELJF	Q9zkb6 helicobacte
754	18	90.0	218	1	RECA_ARTAU	Q9re16 arthrobacte	827	18	90.0	283	1	TVSY_HAEIN	P57808 pasteurella
755	18	90.0	221	1	MLR_DROME	P18432 mycoplasma	828	18	90.0	283	1	TVSY_PASMU	Q25609 helicobacte
756	18	90.0	223	1	DEOC_MYGE	P47296 deoxyplasma	829	18	90.0	284	1	LGT_HELJF	Q8tvt1 methanopyru
757	18	90.0	223	1	SAMP_CAVPO	P49255 cavia porce	830	18	90.0	284	1	NADE_METKA	P95994 sulfobus
758	18	90.0	223	1	SAMP_HUMAN	P02743 homo sapien	831	18	90.0	285	1	Y066_SULSO	P43287 arabidopsis
759	18	90.0	224	1	DEOC_MYCPN	P09924 mycoplasma	832	18	90.0	285	1	PI22_ARATH	P30302 arabidopsis
760	18	90.0	224	1	SAMP_PIG	O19063 sus scrofa	833	18	90.0	286	1	PI11_ARATH	P43285 arabidopsis
761	18	90.0	226	1	ORGB_SALTY	P86654 salmonella	834	18	90.0	286	1	PI12_ARATH	Q06611 arabidopsis
762	18	90.0	228	1	ISPD_ANASP	Q8ylx3 anabaena sp	835	18	90.0	286	1	PI13_ARATH	Q08733 arabidopsis
763	18	90.0	228	1	IPYF_METTH	O26232 methanobact	836	18	90.0	286	1	PI13_ARATH	Q08733 arabidopsis
764	18	90.0	228	1	SAMP_RAT	P23680 rattus norv							

837	18	90.0	286	1	PI25 ARATH	Q9sv31 arabidopsis	910	18	90.0	363	1	LEU3 BUCAP	O85064 buchnera ap
838	18	90.0	286	1	PI25 ARATH	Q8451 lycopersico	911	18	90.0	363	1	MURG ENTFA	O07109 enterococcu
839	18	90.0	287	1	PI14 ARATH	Q39196 arabidopsis	912	18	90.0	364	1	AAT PYRKO	O93744 pyrococcus
840	18	90.0	287	1	PI15 ARATH	Q81a6 arabidopsis	913	18	90.0	364	1	YM25 MYCTU	O10512 mycobacteri
841	18	90.0	287	1	PI21 ARATH	P43286 arabidopsis	914	18	90.0	365	1	HI82 BORBR	O7w4y3 bordetella
842	18	90.0	287	1	PI26 ARATH	Q2v07 arabidopsis	915	18	90.0	365	1	HI82 BORPA	O7w2y3 bordetella
843	18	90.0	289	1	PI27 PEA	Q25794 pisum sativ	916	18	90.0	365	1	HI82 BORPE	O7v8z0 bordetella
844	18	90.0	291	1	EGC1 LISNO	Q8v680 listeria mo	917	18	90.0	365	1	RECA SPIPL	P48293 spirulina p
845	18	90.0	291	1	PI24 ARATH	Q9f553 arabidopsis	918	18	90.0	366	1	IDHA BOVIN	P41563 bos taurus
846	18	90.0	294	1	ISPA BUCAP	Q8k9a0 buchnera ap	919	18	90.0	366	1	IDHA HUMAN	P50213 homo sapien
847	18	90.0	296	1	NI11 ARATH	Q8v2w1 arabidoglob	920	18	90.0	369	1	METE HAEIN	P44502 haemophilus
848	18	90.0	298	1	YK23 ARCFCU	O82w56 archaeoglob	921	18	90.0	369	1	RECA CORPS	P48288 corynebacte
849	18	90.0	299	1	P8D CHLCV	O82113 chlamydomphi	922	18	90.0	369	1	VP6 AHSV6	Q64913 african hor
850	18	90.0	300	1	Y025 NPVOP	O10298 orgyia pseu	923	18	90.0	369	1	VP6 AHSV6	Q64913 african hor
851	18	90.0	302	1	RE04 POPPV	Q9j540 fowlpox vir	924	18	90.0	371	1	Y1B0 CLOAB	O04354 clostridium
852	18	90.0	303	1	EST ACILW	P18773 acinetobact	925	18	90.0	371	1	BIOF METJA	Q58694 methanococc
853	18	90.0	303	1	TRRD HUMAN	Q9nyv9 homo sapien	926	18	90.0	372	1	DEGT BACST	P15263 bacillus st
854	18	90.0	304	1	DCAS AGRTU	Q44185 agrobacteri	927	18	90.0	372	1	RECA CORGL	P42442 corynebacte
855	18	90.0	304	1	KHSE STRAM	Q99ue6 staphylococ	928	18	90.0	376	1	PHNM ECOLI	P16689 escherichia
856	18	90.0	304	1	KHSE STRAM	O8nw8 staphylococ	929	18	90.0	380	1	METC LACLA	Q9r8s9 lactococcus
857	18	90.0	305	1	TRRD RAT	Q9jkt7 rattus norv	930	18	90.0	380	1	YH07 PASMU	O9ckb7 pasteurella
858	18	90.0	306	1	TRUB NEIVA	Q9jtx5 neisseria m	931	18	90.0	382	1	MATB NEUCR	F36981 neurospora
859	18	90.0	307	1	TRUB NEIME	P5j7y1 neisseria m	932	18	90.0	385	1	RNAL SCHPO	O8fbd3 corynebacte
860	18	90.0	310	1	DNUL MYOGE	P47248 mycoplasma	933	18	90.0	385	1	DHAT CIIFR	P41391 schizosacch
861	18	90.0	310	1	YBR1 YEAST	P38238 saccharomyc	934	18	90.0	387	1	GALL FUSNN	Q8rhd0 fusobacteri
862	18	90.0	312	1	ARC2 ENTFA	P59625 enterococcu	935	18	90.0	389	1	Y4QE RHISN	P55626 rhizobium s
863	18	90.0	312	1	V101 VACOC	P20498 vaccinia vi	936	18	90.0	393	1	CYS3 YEAST	P31373 saccharomyc
864	18	90.0	312	1	V101 VAVV	P32999 variola vir	937	18	90.0	393	1	HEM2 YEAST	P16622 saccharomyc
865	18	90.0	312	1	PEPM STRVR	O86937 streptomyc	938	18	90.0	394	1	AAT AQUAE	O67781 aquifex aeo
866	18	90.0	313	1	Y376 AQUAE	O66638 aquifex aeo	939	18	90.0	394	1	VORA PYRPU	O51801 pyrococcus
867	18	90.0	315	1	O3A1 HUMAN	P47881 homo sapien	940	18	90.0	394	1	VORA PYRPU	O51801 pyrococcus
868	18	90.0	315	1	O3A1 PANTR	Q9tua4 pan troglod	941	18	90.0	394	1	VORA PYRAB	O9uy21 pyrococcus
869	18	90.0	316	1	DHAS VIBMI	Q60080 vibrio mimi	942	18	90.0	395	1	PGK STRPB	O8293 streptococc
870	18	90.0	316	1	Y151 PYRPU	Q8t2x4 pyrococcus	943	18	90.0	397	1	PGK STRPB	O8293 streptococc
871	18	90.0	318	1	Y211 AQUAE	O66405 aquifex aeo	944	18	90.0	397	1	PGK STRPB	O8293 streptococc
872	18	90.0	321	1	LPXD CMJUE	Q9phu0 campylobact	945	18	90.0	401	1	ASSY AQUAE	O67213 aquifex aeo
873	18	90.0	322	1	YQTA BACSU	P54538 bacillus su	946	18	90.0	401	1	HISX ARCFCU	O30027 archaeoglob
874	18	90.0	326	1	YQTA ARATH	P52839 arabidopsis	947	18	90.0	404	1	ISCS BUCBP	O89a19 buchnera ap
875	18	90.0	330	1	LDHD STRA3	Q8e8a9 streptococc	948	18	90.0	405	1	ENTM BACFR	P54355 bacteroides
876	18	90.0	330	1	LDHD STRA3	Q8e0n5 streptococc	949	18	90.0	405	1	Y872 METJA	O58282 methanococc
877	18	90.0	330	1	LDHD STRA3	Q99m2 streptococc	950	18	90.0	406	1	DEOB AGRT5	O58282 methanococc
878	18	90.0	333	1	DHD2 LACPA	P17584 lactobacilli	951	18	90.0	406	1	DEOB RHIME	O57000 mycobacteri
879	18	90.0	336	1	PEXG HUMAN	Q9v5y5 homo sapien	952	18	90.0	408	1	YFIN ECOLI	O28596 archaeoglob
880	18	90.0	336	1	RUL MCPE	Q8ex24 mycoplasma	953	18	90.0	408	1	YLO1 SCHPO	O28596 archaeoglob
881	18	90.0	337	1	DHAS VIBCH	P23247 vibrio chol	954	18	90.0	410	1	YV28 MYCTU	O8uJ04 agrobacteri
882	18	90.0	339	1	DUS RICCN	Q92jg6 rickettsia	955	18	90.0	411	1	TRUD ARCFCU	Q92t47 rhizobium m
883	18	90.0	340	1	ASTG PSEAE	P80358 pseudomonas	956	18	90.0	414	1	GAG2 DROME	P46139 escherichia
884	18	90.0	340	1	LINI YEAST	P38852 saccharomyc	957	18	90.0	414	1	ODO2 ALCEU	O13883 schizosacch
885	18	90.0	340	1	TRPD AQUAE	O65576 aquifex aeo	958	18	90.0	416	1	GLYA ACIRA	O50700 mycobacteri
886	18	90.0	341	1	DUS RCPR	Q9zed3 rickettsia	959	18	90.0	417	1	YB09 YEAST	O28596 archaeoglob
887	18	90.0	342	1	ARSA CAEL	P30632 caenorhabdi	960	18	90.0	420	1	ESC DROME	O28596 archaeoglob
888	18	90.0	342	1	YQ66 CAEL	Q9275 caenorhabdi	961	18	90.0	425	1	ESC DROME	O28596 archaeoglob
889	18	90.0	343	1	HRCQ THETN	Q8rb70 thermotanaer	962	18	90.0	425	1	NOOF AQUAE	O26458 drosophila
890	18	90.0	343	1	MJK2 METJA	O88752 methanococc	963	18	90.0	426	1	NOOF AQUAE	O26458 drosophila
891	18	90.0	344	1	ABTC LACIA	Q01457 lactococcus	964	18	90.0	429	1	GLYA METJA	O58932 methanococc
892	18	90.0	345	1	YF46 METJA	Q58941 methanococc	965	18	90.0	430	1	SPAL SHIFL	P35531 shigella fl
893	18	90.0	347	1	IDHA VACPA	Q28480 macaca fasc	966	18	90.0	433	1	SSAN SALTU	P74857 salmonella
894	18	90.0	348	1	ARSI HUMAN	Q43681 homo sapien	967	18	90.0	433	1	INCK ECOLI	P22937 escherichia
895	18	90.0	348	1	ARSI MOUSE	O43984 mus musculu	968	18	90.0	434	1	PEL LILLO	P40973 lilium long
896	18	90.0	348	1	GTOM ARATH	Q9zrk1 arabidopsis	969	18	90.0	438	1	TRME URSPA	Q9prc7 ureaplasma
897	18	90.0	348	1	ODPA RHIME	Q9zrk1 rhizobium m	970	18	90.0	440	1	MANA CANAL	P34948 candida alb
898	18	90.0	348	1	RECA SYN2	P44582 synecococc	971	18	90.0	444	1	YBWA BACSU	P39584 bacillus su
899	18	90.0	352	1	ARGC SULSO	Q980x1 sulfolobus	972	18	90.0	446	1	SUB2 YEAST	Q07478 saccharomyc
900	18	90.0	352	1	YPBB BACSU	P30728 bacillus su	973	18	90.0	451	1	RADA TREPA	P38295 saccharomyc
901	18	90.0	353	1	RECA CHLEN	Q9z7e4 chlamydia p	974	18	90.0	455	1	PURB BUCAI	O83985 treponema p
902	18	90.0	354	1	RECA SYN3	P74737 synecocyst	975	18	90.0	456	1	PURB BUCAI	P57351 buchnera ap
903	18	90.0	355	1	RECA SYNEL	Q8dh70 synecococc	976	18	90.0	460	1	UDPE GVLO	Q38166 lachnobia o
904	18	90.0	357	1	PMFE PROMI	P53522 proteus mir	977	18	90.0	462	1	YB47 METJA	Q58547 methanococc
905	18	90.0	357	1	RECA ANASP	P58552 anabaena ep	978	18	90.0	465	1	YHJA ECOLI	P37197 escherichia
906	18	90.0	358	1	RECA ANAVA	P14167 anabaena va	979	18	90.0	467	1	ARLY RANCA	P51464 rana catesb
907	18	90.0	358	1	RNFD HAEIN	Q57288 haemophilus	980	18	90.0	469	1	SELB METJA	Q57918 methanococc
908	18	90.0	359	1	RPBG SALTU	P46397 salmonella	981	18	90.0	473	1	ATPB RICCN	Q92988 rickettsia
909	18	90.0	360	1	BUX ENTFA	Q9rps7 enterococcu	982	18	90.0	473	1	NIFD_FRAAL	Q02452 frankia aln

883 18 50.0 474 1 SVC PYRAE  
 884 18 50.0 477 1 YGJI ECOLI  
 885 18 50.0 484 1 GUNA XANCP  
 886 18 50.0 490 1 DCLY BACSU  
 887 18 50.0 498 1 SECY MYGCA  
 888 18 50.0 499 1 SVFA THACA  
 889 18 50.0 500 1 GALT CAEEL  
 890 18 50.0 500 1 N04C ARATH  
 891 18 50.0 506 1 PUR9 ANASP  
 892 18 50.0 507 1 C4DE DROME  
 893 18 50.0 509 1 ATPA RHACA  
 894 18 50.0 511 1 Y538 YEAST  
 895 18 50.0 513 1 Y538 RICPR  
 896 18 50.0 529 1 PUR9 CAUCR  
 897 18 50.0 529 1 VSM6 TRTBB  
 898 18 50.0 529 1 YB89 YEAST  
 899 18 50.0 532 1 FMO5 CAVPO  
 1000 18 50.0 536 1 ENTE\_ECO57

## ALIGNMENTS

RESULT 1  
 ID NS2 MYCTU STANDARD; PRT; 19 AA.  
 AC P81136;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 30 kDa non-secretory protein 2 (fragment).  
 DS Mycobacterium tuberculosis.  
 DC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 DC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 DX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=H37Rv;  
 RA Prasad H.K., Annapurna P.S.;  
 RL Submitted (DEC-1997) to Swiss-Prot.  
 SC -1- SIMILARITY: STRONG, TO THE N-TERMINALS OF C.ELEGANS F20A1.4  
 SC AND H. INFLUENZAE H10967.  
 SC -1- CAUTION: We are unable to find this protein in the translation of  
 SC the genome of strain H37Rv.  
 ST NON TER 19  
 SQ SEQUENCE 19 AA; 2211 MW; ABC1854BFF1FF70 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 11 VAEF 14

RESULT 2  
 SECE TREPA STANDARD; PRT; 59 AA.  
 AC O83263;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Preprotein translocase secE subunit.  
 DN SECE OR TP0235.  
 DS Treponema pallidum.  
 DC Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.  
 DX NCBI\_TaxID=160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Nichols;  
 RX MEDLINE=98332770; PubMed=9668876;  
 RA Fraser C.N., Norris S.J., Weinstock G.M., White O., Sutton G.G.,

RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,  
 RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete.";  
 RL Science 281:375-388(1998).  
 CC -1- FUNCTION: Essential for protein export.  
 CC -1- SUBCELLULAR LOCATION: Tail-anchored membrane protein (Potential).  
 CC -1- SIMILARITY: Belongs to the secE/SEC61-gamma family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE001205; AAC65223.1; -;  
 DR PIR; E71349; E71349.  
 DR TIGR; TP0235; -;  
 DR InterPro; IPR001901; SecE.  
 DR InterPro; IPR005807; SecE\_bac.  
 DR Pfam; PF00584; SecE; 1.  
 DR TIGRFAMs; TIGR00964; aa0501s06; 1.  
 DR PROSITE; PS01067; SECE\_SEC61G; 1.  
 KW Protein transport; Translocation; Transmembrane; Complete proteome.  
 SQ TRANSEM 39 59 POTENTIAL.  
 FT SEQUENCE 59 AA; 6789 MW; 9AC35BA8F4B2A7F CRC64;

Query Match 95.0%; Score 19; DB 1; Length 59;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 12 VAEF 15

RESULT 3  
 YA90 ARCFU STANDARD; PRT; 59 AA.  
 ID YA90 ARCFU  
 AC O29175;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Hypothetical UPF0165 protein AFI090.  
 GN AFI090.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
 OC Archaeoglobaceae; Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,  
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Ardea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "the complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).

```

-!- SIMILARITY: Belongs to the UPF0165 family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE001028; AAC90157.1; -.
TIGR; A69386; A69386.
InterPro; IPR008203; DUF104.
InterPro; IPR008204; DUF104_N.
Pfam; PF01954; DUF104; 1.
ProDom; PD005964; DUF104_N; 1.
Hypothetical protein; Complete proteome.
SEQUENCE 59 AA; 6887 MW; B72B847382B2B6C6 CRC64;
-----
Query Match 95.0%; Score 19; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
2 VAEF 5
|||||
51 VAEF 54
-----
RESULT 4
SI_HAEMIN STANDARD; PRT; 63 AA.
P43998;
01-NOV-1995 (Rel. 32, Created)
01-NOV-1995 (Rel. 32, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Hypothetical protein HI0451.
Haemophilus influenzae.
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Haemophilus.
NCBI_TaxID=727;
[1]
SEQUENCE FROM N.A.
STRAIN-Rd / KW20 / ATCC 51907;
MEDLINE=95350630; PubMed=7542800;
Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
Fine L.D., Fritchman J.L., Fuhrmann J.B., Geoghagen N.S.M.,
Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
Venter J.C.;
"Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
Science 269:496-512(1995).
-!- FUNCTION: This protein is one of the 16S ribosomal RNA binding
proteins (By similarity).
-!- SIMILARITY: Belongs to the S15P family of ribosomal proteins.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U32812; AAC22973.1; -.
EMBL; U32825; AAC23117.1; -.
PIR; H64116; H64116.
HSSP; P05766; 1A32.
TIGR; HI1328; -.
TIGR; HI1468; -.
DR InterPro; IPR000589; Ribosomal_S15.
DR InterPro; IPR005290; Ribosomal_S15_b.
DR Pfam; PF00312; Ribosomal_S15; 1.
DR ProDom; PD157043; RS15_bact; 1.
DR TIGRFAMs; TIGR00952; S15_bact; 1.
DR PROSITE; PS00362; RIBOSOMAL_S15; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
FT INIT MET 0 BY SIMILARITY.
SQ SEQUENCE 88 AA; 10064 MW; 563BAD2B8B8A7043 CRC64;
-----
Query Match 95.0%; Score 19; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
2 VAEF 5
|||||
11 VAEF 14
-----

```

```

Query Match 95.0%; Score 19; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
QY 2 VAEF 5
DB 27 VAEF 30
-----
RESULT 5
RS15_HAEMIN STANDARD; PRT; 88 AA.
ID RS15_HAEMIN
AC P43389;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S15.
DE (RPSO-A OR RPS15-A OR HI1328) AND (RPSO-B OR RPS15-B OR HI1468).
GN Haemophilus influenzae.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.B., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: This protein is one of the 16S ribosomal RNA binding
proteins (By similarity).
CC -!- SIMILARITY: Belongs to the S15P family of ribosomal proteins.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U32812; AAC22973.1; -.
EMBL; U32825; AAC23117.1; -.
PIR; H64116; H64116.
HSSP; P05766; 1A32.
TIGR; HI1328; -.
TIGR; HI1468; -.
DR InterPro; IPR000589; Ribosomal_S15.
DR InterPro; IPR005290; Ribosomal_S15_b.
DR Pfam; PF00312; Ribosomal_S15; 1.
DR ProDom; PD157043; RS15_bact; 1.
DR TIGRFAMs; TIGR00952; S15_bact; 1.
DR PROSITE; PS00362; RIBOSOMAL_S15; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
FT INIT MET 0 BY SIMILARITY.
SQ SEQUENCE 88 AA; 10064 MW; 563BAD2B8B8A7043 CRC64;
-----
Query Match 95.0%; Score 19; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
QY 2 VAEF 5
DB 11 VAEF 14
-----

```

OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=9637999; PubMed=868087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geobach N.S.M., Weidman J.P., Fuhrmann J.L., Nguyen D.,  
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii";  
RL Science 273:1058-1073(1996).  
CC -!- SIMILARITY: BELONGS TO THE UPF0166 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U67593; AAB99550.1; -;  
DR PIR; C64490; C64490.  
DR TIGR; MJ1524; -;  
DR InterPro; IPR003793; DUF190.  
DR Pfam; PF02641; DUF190; 1.  
DR Hypothetical protein; Complete proteome.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 108 AA; 12398 MW; D3E286AFB9D4AF8B CRC64;  
  
Query Match 95.0%; Score 19; DB 1; Length 108;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 VAEF 5  
DB 52 VAEF 55  
  
RESULT 8  
KEDA ACTSL STANDARD; PRT; 114 AA.  
AC P41249;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Apokedarcidin.  
OS Actinomycete sp. (strain L585-6 / ATCC 53650).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales.  
OX NCBI\_TaxID=38989;  
RN [1]  
RP SEQUENCE.  
RC STRAIN=ATCC 53650 / L585-6;  
RX MEDLINE=93015257; PubMed=1399845;  
RA Hofstead S.J., Watson J.A., Malacko A.R., Marquardt H.;  
RT "Kedarcidin, a new chromoprotein antitumor antibiotic. II. Isolation,  
RT purification and physico-chemical properties.";  
RL J. Antibiot. 45:1250-1254(1992).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=93376732; PubMed=8367457;  
RA Zein N., Casazza A.M., Doyle T.W., Leet J.E., Scheoeder D.R.,  
RA Solomon W., Nadler S.G.;  
RT "Selective proteolytic activity of the antitumor agent kedarcidin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:8009-8012(1993).  
RN [3]

OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=9637999; PubMed=868087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geobach N.S.M., Weidman J.P., Fuhrmann J.L., Nguyen D.,  
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii";  
RL Science 273:1058-1073(1996).  
CC -!- SIMILARITY: BELONGS TO THE UPF0166 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U67593; AAB99550.1; -;  
DR PIR; C64490; C64490.  
DR TIGR; MJ1524; -;  
DR InterPro; IPR003793; DUF190.  
DR Pfam; PF02641; DUF190; 1.  
DR Hypothetical protein; Complete proteome.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 108 AA; 12398 MW; D3E286AFB9D4AF8B CRC64;  
  
Query Match 95.0%; Score 19; DB 1; Length 108;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 VAEF 5  
DB 52 VAEF 55  
  
RESULT 8  
KEDA ACTSL STANDARD; PRT; 114 AA.  
AC P41249;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Apokedarcidin.  
OS Actinomycete sp. (strain L585-6 / ATCC 53650).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales.  
OX NCBI\_TaxID=38989;  
RN [1]  
RP SEQUENCE.  
RC STRAIN=ATCC 53650 / L585-6;  
RX MEDLINE=93015257; PubMed=1399845;  
RA Hofstead S.J., Watson J.A., Malacko A.R., Marquardt H.;  
RT "Kedarcidin, a new chromoprotein antitumor antibiotic. II. Isolation,  
RT purification and physico-chemical properties.";  
RL J. Antibiot. 45:1250-1254(1992).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=93376732; PubMed=8367457;  
RA Zein N., Casazza A.M., Doyle T.W., Leet J.E., Scheoeder D.R.,  
RA Solomon W., Nadler S.G.;  
RT "Selective proteolytic activity of the antitumor agent kedarcidin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:8009-8012(1993).  
RN [3]

OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=9637999; PubMed=868087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geobach N.S.M., Weidman J.P., Fuhrmann J.L., Nguyen D.,  
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii";  
RL Science 273:1058-1073(1996).  
CC -!- SIMILARITY: BELONGS TO THE UPF0166 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U67593; AAB99550.1; -;  
DR PIR; C64490; C64490.  
DR TIGR; MJ1524; -;  
DR InterPro; IPR003793; DUF190.  
DR Pfam; PF02641; DUF190; 1.  
DR Hypothetical protein; Complete proteome.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 108 AA; 12398 MW; D3E286AFB9D4AF8B CRC64;  
  
Query Match 95.0%; Score 19; DB 1; Length 93;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 VAEF 5  
DB 47 VAEF 50  
  
RESULT 7  
KEDA ACTSL STANDARD; PRT; 108 AA.  
AC P41249;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Apokedarcidin.  
OS Actinomycete sp. (strain L585-6 / ATCC 53650).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales.  
OX NCBI\_TaxID=38989;  
RN [1]  
RP SEQUENCE.  
RC STRAIN=ATCC 53650 / L585-6;  
RX MEDLINE=93015257; PubMed=1399845;  
RA Hofstead S.J., Watson J.A., Malacko A.R., Marquardt H.;  
RT "Kedarcidin, a new chromoprotein antitumor antibiotic. II. Isolation,  
RT purification and physico-chemical properties.";  
RL J. Antibiot. 45:1250-1254(1992).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=93376732; PubMed=8367457;  
RA Zein N., Casazza A.M., Doyle T.W., Leet J.E., Scheoeder D.R.,  
RA Solomon W., Nadler S.G.;  
RT "Selective proteolytic activity of the antitumor agent kedarcidin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:8009-8012(1993).  
RN [3]

OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 430

```

2  STRUCTURE BY NMR.
3  STRAIN=ATCC 53650 / L585-6;
4  MEDLINE=95001848; PubMed=7918358;
5  Constantine K.L., Colson K.L., Wittekind M., Friedrichs M.S.,
6  Zein N., Tuttle J., Langley D.R., Leet J.E., Schroeder D.R., Lam K.S.,
7  Farmer B.T. II, Metzler W.J., Bruccoleri R.E., Mueller L.,
8  "Sequential 1H, 13C, and 15N NMR assignments and solution
9  conformation of apokardacin.";
10 Biochemistry 33:11438-11452(1994).
11 -!- FUNCTION: BINDS NON-COVALENTLY TO AN ENEDIYNE CHROMOPHORE WHICH IS
12 THE CYTOTOXIC AND MUTAGENIC COMPONENT OF THE ANTIBIOTIC. THE
13 CHROMOPHORE CLEAVES DUPLEX DNA SITE-SPECIFICALLY IN A SINGLE-
14 STRANDED MANNER. THE APOPROTEIN CLEAVES PROTEINS SELECTIVELY IN
15 PARTICULAR HIGHLY BASIC HISTONES, WITH H1 PROTEINS BEING CLEAVED
16 THE MORE READILY.
17 -!- DOMAIN: THIS PROTEIN CONSISTS OF AN IMMUNOGLOBULIN-LIKE SEVEN-
18 STRANDED ANTIPARALLEL BETA-BARREL DOMAIN LINKED TO A SUBDOMAIN
19 COMPOSED OF TWO BETA-HAIRPIN RIBBONS.
20 -!- SIMILARITY: Belongs to the neocarzinostatin family.
21 PDB: 1AKP; 31-AUG-94.
22 InterPro; IPR002186; Neocarzinostat.
23 Pfam; PF00960; Neocarzinostat; 1.
24 ProDom; PD012709; Neocarzinostat; 1.
25 Antibiotic; DNA-binding; 3D-structure.
26 DISULFID 37 47
27 STRAND 88 95
28 STRAND 4 7
29 STRAND 11 13
30 TURN 15 16
31 STRAND 15 18
32 STRAND 21 24
33 STRAND 32 33
34 STRAND 36 39
35 STRAND 46 47
36 STRAND 50 51
37 TURN 55 55
38 STRAND 61 61
39 STRAND 64 64
40 STRAND 67 67
41 STRAND 71 74
42 TURN 77 79
43 STRAND 84 87
44 STRAND 94 97
45 STRAND 100 100
46 STRAND 109 110
47 STRAND 112 114
48 SEQUENCE 114 AA; 10969 MW; 1901E2B14E4197B4 CRC64;
49
50 Query Match 95.0%; Score 19; DB 1; Length 114;
51 Best Local Similarity 100.0%; Pred. No. 2e+02;
52 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
53
54 2 VAEF 5
55 ||||
56 49 VAEF 52
57
58 RESULT 9
59 IS VERMO STANDARD; PRT; 115 AA.
60 Q9W7R2;
61 16-OCT-2001 (Rel. 40, Created)
62 16-OCT-2001 (Rel. 40, Last sequence update)
63 10-OCT-2003 (Rel. 42, Last annotation update)
64 Insulin precursor.
65 INS.
66 Verasper moseri (Barfin flounder).
67 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
68 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
69 Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
70 Pleuronectoidae; Pleuronectidae; Verasper.
71 NCBI_TaxID=98923;
72 [1]
73
74 SEQUENCE FROM N.A.
75 Andoh T., Nagasawa H.;
76 "Two molecular forms of insulin from barfin flounder, Verasper moseri,
77 are derived from a single gene.";
78 Zool. Sci. 15:931-937(1998).
79 -!- FUNCTION: Insulin decreases blood glucose concentration. It
80 increases cell permeability to monosaccharides, amino acids and
81 fatty acids. It accelerates glycolysis, the pentose phosphate
82 cycle, and glycogen synthesis in liver.
83 -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
84 disulfide bonds.
85 -!- SUBCELLULAR LOCATION: Secreted.
86 -!- SIMILARITY: Belongs to the insulin family.
87
88 This SWISS-PROT entry is copyright. It is produced through a collaboration
89 between the Swiss Institute of Bioinformatics and the EMBL outstation
90 the European Bioinformatics Institute. There are no restrictions on its
91 use by non-profit institutions as long as its content is in no way
92 modified and this statement is not removed. Usage by and for commercial
93 entities requires a license agreement (See http://www.isb-sib.ch/announce/
94 or send an email to license@isb-sib.ch).
95
96 EMBL; AB029318; BAA82315.1; -
97 HSSP; P01315; 1MPJ.
98 InterPro; IPR004825; Ins/IGF/relax.
99 Pfam; PF00049; Insulin; 1.
100 PRINTS; PS00277; INSULINB.
101 SMART; SM00078; ILGF; 1.
102 PROSITE; PS00262; INSULIN; 1.
103 Insulin family; Hormone; Glucose metabolism; Signal.
104 SIGNAL 1 22 BY SIMILARITY.
105 FT CHAIN 23 53 INSULIN B CHAIN.
106 FT PROPEP 56 92 C PEPTIDE.
107 FT CHAIN 95 115 INSULIN A CHAIN.
108 FT DISULFID 32 101 INTERCHAIN (BY SIMILARITY).
109 FT DISULFID 44 114 INTERCHAIN (BY SIMILARITY).
110 FT DISULFID 100 105 BY SIMILARITY.
111 SQ SEQUENCE 115 AA; 12608 MW; 7EA2A5B56DEDDDBB CRC64;
112
113 Query Match 95.0%; Score 19; DB 1; Length 115;
114 Best Local Similarity 100.0%; Pred. No. 2e+02;
115 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
116
117 QY 2 VAEF 5
118 ||||
119 Db 79 VAEF 82
120
121 RESULT 10
122 INS_LOPPI STANDARD; PRT; 116 AA.
123 AC P01341;
124 DT 21-JUL-1986 (Rel. 01, Created)
125 DT 21-JUL-1986 (Rel. 01, Last sequence update)
126 DT 10-OCT-2003 (Rel. 42, Last annotation update)
127 DE Insulin precursor.
128 GN INS.
129 OS Lophius piscatorius (Allmouth goosefish) (Anglerfish), and
130 Lophius americanus (American goosefish) (Anglerfish).
131 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
132 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
133 OC Acanthomorpha; Paracanthopterygii; Lophiiformes; Lophiidae; Lophius.
134 OC NCBI_TaxID=8074, 8073;
135 RN [1]
136 RP SEQUENCE FROM N.A.
137 RC SPECIES=L.americianus;
138 RX MEDLINE=81056434; PubMed=7001633;
139 RA Hobart P.M., Shen L.-P., Crawford R., Pictet R.L., Rutter W.J.;
140 "Comparison of the nucleic acid sequence of anglerfish and mammalian
141 insulin mRNAs from cloned cDNA's.";
142 Science 210:1360-1363(1980).
143 RL [2]
144 RP SEQUENCE OF 25-54 AND 96-116.

```

```

RC SPCIBS=L.piscatorius;
RX MEDLINE=70036620; PubMed=539299;
RA Neumann P.A., Koldenhu M., Humbel R.E.;
RT "Amino acid sequence of insulin from the angler fish (Ioplius
RT piscatorius).";
UL Hoppe-Seyler's Z. Physiol. Chem. 350:1286-1288(1969).
XC -!- FUNCTION: Insulin decreases blood glucose concentration. It
XC increases cell permeability to monosaccharides, amino acids and
XC fatty acids. It accelerates glycolysis, the pentose phosphate
XC cycle, and glycogen synthesis in liver.
XC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
XC disulfide bonds.
XC -!- SUBCELLULAR LOCATION: Secreted.
XC -!- SIMILARITY: Belongs to the insulin family.
XC
XC This SWISS-PROT entry is copyright. It is produced through a collaboration
XC between the Swiss Institute of Bioinformatics and the EMBL outstation -
XC the European Bioinformatics Institute. There are no restrictions on its
XC use by non-profit institutions as long as its content is in no way
XC modified and this statement is not removed. Usage by and for commercial
XC entities requires a license agreement (see http://www.isb-sib.ch/announce/
XC or send an email to license@isb-sib.ch).
XC
XC EMBL; V00634; CAA23907.1; -.
RX PIR; A01608; IPAF.
RX HSSP; P01308; ILPH.
RX InterPro; IPR004825; Ins/IGF/relax.
RX Pfam; PF00049; Insulin; 1.
RX PRINTS; PR00277; INSULINB.
RX SMART; SM00078; IIGF; 1.
RX PROSITE; PS00262; INSULIN; 1.
RX Insulin family; Hormone; Glucose metabolism; Signal.
RX SIGNAL 1
RX CHAIN 25 53 INSULIN B CHAIN.
RX PROPEP 56 93 C PEPTIDE.
RX CHAIN 96 116 INSULIN A CHAIN.
RX DISULFID 32 102 INTERCHAIN.
RX DISULFID 44 115 INTERCHAIN.
RX DISULFID 101 106
RX SEQUENCE 116 AA; 12737 MW; C686F8F81B3BEFE CRC64;
RX
RX Query Match 95.0%; Score 19; DB 1; Length 116;
RX Best Local Similarity 100.0%; Pred. No. 2.1e+02;
RX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
RX
RX 2 VAEF 5
RX 80 VAEF 83
RX
RX RESULT 11
RX U279 DROME STANDARD; PRT; 119 AA.
RX AC Q9V8F3;
RX DT 10-OCT-2003 (Rel. 42, Created)
RX DT 10-OCT-2003 (Rel. 42, Last sequence update)
RX DT 10-OCT-2003 (Rel. 42, Last annotation update)
RX DE UPG0279 protein CG14505.
RX GN CG14505.
RX OS Drosophila melanogaster (Fruit fly).
RX OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
RX OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
RX OC Ephydroidea; Drosophilidae; Drosophila.
RX OC NCBI_TaxID=7227;
RX RN [1]
RX RP SEQUENCE FROM N.A.
RX RC STRAIN=Berkley;
RX RX MEDLINE=20196006; PubMed=10731132;
RX RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RX RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RX RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RX RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RX RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

```

```

RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.P., Achayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.F., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,
RA Burtis J.M., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.C., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evanselista C.C., Ferraz C., Perriera S., Fleischmann W.,
RA Fiedler C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Glöckner A., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RA Science 287:2185-2195(2000).
RA [2]
RA SEQUENCE FROM N.A.
RX STRAIN=Berkley;
RX MEDLINE=22426066; PubMed=12537569;
RX Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RX George R.A., Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,
RX Rubin G.M., Celnik S.E.;
RX "A Drosophila full-length cDNA resource."
RX Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RX -!- SIMILARITY: Belongs to the UPG0279 family.
RX
RX This SWISS-PROT entry is copyright. It is produced through a collaboration
RX between the Swiss Institute of Bioinformatics and the EMBL outstation -
RX the European Bioinformatics Institute. There are no restrictions on its
RX use by non-profit institutions as long as its content is in no way
RX modified and this statement is not removed. Usage by and for commercial
RX entities requires a license agreement (see http://www.isb-sib.ch/announce/
RX or send an email to license@isb-sib.ch).
RX
RX EMBL; AE003800; AAF57714.1; -.
RX EMBL; AY084100; AAL89838.1; -.
RX FlyBase; FBgn0034327; CGI4505.
RX InterPro; IPR007967; DUF727.
RX Pfam; PF05303; DUF727; 1.
RX SEQUENCE 119 AA; 13534 MW; 83FA23FCCCE389AA CRC64;
RX
RX Query Match 95.0%; Score 19; DB 1; Length 119;
RX Best Local Similarity 100.0%; Pred. No. 2.1e+02;
RX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
RX
RX 2 VAEF 5
RX 33 VAEF 36
RX
RX RESULT 12
RX GTR2_PIG STANDARD; PRT; 120 AA.
RX ID GTR2_PIG
RX AC O62786;
RX DT 28-FEB-2003 (Rel. 41, Created)

```

```
28-FEB-2003 (Rel. 41, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Solute carrier family 2, facilitated glucose transporter, member 2
(Glucose transporter type 2, liver) (Fragment).
SLC2A2 OR GLUT2.
Sus scrofa (Pig).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxID=9823;
[1]
SEQUENCE FROM N.A.
Canty J.M., Young R.F., Fallavollita J.A.;
Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: Facilitative glucose transporter. This isoform likely
mediates the bidirectional transfer of glucose across the plasma
membrane of hepatocytes and is responsible for uptake of glucose
by the beta cells; may comprise part of the glucose-sensing
mechanism of the beta cell. May also participate with the
Na(+)/glucose cotransporter in the transcellular transport of
glucose in the small intestine and kidney (By similarity).
-!- SUBCELLULAR LOCATION: Integral membrane protein.
-!- SIMILARITY: Belongs to the sugar transporter family. Glucose
transporter subfamily.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF054835; AAC12737.1; -.
InterPro; IPR007114; MFS.
InterPro; IPR005828; Sub transporter.
InterPro; IPR005829; Sug transporter.
InterPro; IPR003663; Sugar transp.
Pfam; PF00083; sugar tr; 1
PRINTS; PR00171; SUGSTRANSPT.
PROSITE; PS00850; MFS; 1.
PROSITE; PS00216; SUGAR_TRANSPORT_1; PARTIAL.
PROSITE; PS00217; SUGAR_TRANSPORT_2; PARTIAL.
Transmembrane; Sugar transport; Transp; Multigene family.
NON_TER 1 1
TRANSMEM 2 22 9 (POTENTIAL).
DOMAIN 23 26 EXTRACELLULAR (POTENTIAL).
TRANSMEM 27 47 10 (POTENTIAL).
DOMAIN 48 56 CYTOPLASMIC (POTENTIAL).
TRANSMEM 57 77 11 (POTENTIAL).
DOMAIN 78 84 EXTRACELLULAR (POTENTIAL).
TRANSMEM 85 105 12 (POTENTIAL).
DOMAIN 106 120 CYTOPLASMIC (POTENTIAL).
NON_TER 120 120
SEQUENCE 120 AA; D5F73168BDF03203 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
|
|
|
O 46 VAEF 49

RESULT 13
18E_PYRAE STANDARD; PRT; 122 AA.
D R18E_PYRAE
C Q8ZY02;
T 28-FEB-2003 (Rel. 41, Created)
T 28-FEB-2003 (Rel. 41, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
E S0S ribosomal protein L18e.
N RPL18E OR PAB0672.

28-FEB-2003 (Rel. 41, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Solute carrier family 2, facilitated glucose transporter, member 2
(Glucose transporter type 2, liver) (Fragment).
SLC2A2 OR GLUT2.
Sus scrofa (Pig).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxID=9823;
[1]
SEQUENCE FROM N.A.
Canty J.M., Young R.F., Fallavollita J.A.;
Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: Facilitative glucose transporter. This isoform likely
mediates the bidirectional transfer of glucose across the plasma
membrane of hepatocytes and is responsible for uptake of glucose
by the beta cells; may comprise part of the glucose-sensing
mechanism of the beta cell. May also participate with the
Na(+)/glucose cotransporter in the transcellular transport of
glucose in the small intestine and kidney (By similarity).
-!- SUBCELLULAR LOCATION: Integral membrane protein.
-!- SIMILARITY: Belongs to the sugar transporter family. Glucose
transporter subfamily.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF054835; AAC12737.1; -.
InterPro; IPR007114; MFS.
InterPro; IPR005828; Sub transporter.
InterPro; IPR005829; Sug transporter.
InterPro; IPR003663; Sugar transp.
Pfam; PF00083; sugar tr; 1
PRINTS; PR00171; SUGSTRANSPT.
PROSITE; PS00850; MFS; 1.
PROSITE; PS00216; SUGAR_TRANSPORT_1; PARTIAL.
PROSITE; PS00217; SUGAR_TRANSPORT_2; PARTIAL.
Transmembrane; Sugar transport; Transp; Multigene family.
NON_TER 1 1
TRANSMEM 2 22 9 (POTENTIAL).
DOMAIN 23 26 EXTRACELLULAR (POTENTIAL).
TRANSMEM 27 47 10 (POTENTIAL).
DOMAIN 48 56 CYTOPLASMIC (POTENTIAL).
TRANSMEM 57 77 11 (POTENTIAL).
DOMAIN 78 84 EXTRACELLULAR (POTENTIAL).
TRANSMEM 85 105 12 (POTENTIAL).
DOMAIN 106 120 CYTOPLASMIC (POTENTIAL).
NON_TER 120 120
SEQUENCE 120 AA; D5F73168BDF03203 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
|
|
|
O 46 VAEF 49

RESULT 13
18E_PYRAE STANDARD; PRT; 122 AA.
D R18E_PYRAE
C Q8ZY02;
T 28-FEB-2003 (Rel. 41, Created)
T 28-FEB-2003 (Rel. 41, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
E S0S ribosomal protein L18e.
N RPL18E OR PAB0672.

Pyrobaculum aerophilum.
Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
Thermoproteaceae; Pyrobaculum.
NCBI_TaxID=13773;
[1]
SEQUENCE FROM N.A.
STRAIN=IM2 / ATCC 51768 / DSM 7523;
MEDLINE=21664397; PubMed=11792889;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
aerophilum.";
Proc. Natl. Acad. Sci. U.S.A. 99; 984-989 (2002).
-!- SIMILARITY: Belongs to the L18E family of ribosomal proteins.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE009779; AAL62941.1; -.
HAMAP; MF_00329; -.
InterPro; IPR001196; Ribosomal L15.
InterPro; IPR000039; Ribosomal_L18e.
Pfam; PF0256; L15; 1.
PROSITE; PS01106; RIBOSOMAL_L18E; FALSE NEG.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 122 AA; 13252 MW; 88F3DB732C4E0394 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 122;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
|
|
|
DB 35 VAEF 38

RESULT 14
Y670_PASMU STANDARD; PRT; 124 AA.
ID Y670_PASMU
AC Q9CMY0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein PM0670 precursor.
GN PM0670.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OC NCBI_TaxID=747;
RN [1]
SEQUENCE FROM N.A.
STRAIN=Pm70;
MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
Proc. Natl. Acad. Sci. U.S.A. 98; 3460-3465 (2001).
-!- SIMILARITY: Belongs to the cytochrome b562 family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE006103; AAK02754.1; -.
PIRSF; PIRSF000029; Cytochrome_b562; 1.
```

```

W Hypothetical protein; Signal; Complete proteome.
T SIGNAL 1 23 POTENTIAL.
T CHAIN 24 124 HYPOTHETICAL PROTEIN PM0670.
Q SEQUENCE 124 AA; 13746 MW; D7B2B485C7B51B9A CRC64;

Query Match 95.0%; Score 19; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
b 45 VAEF 48

RESULT 15
LBP_MOUSE STANDARD; PRT; 127 AA.
D - ILBP_MOUSE
C PS1162;
T 01-OCT-1996 (Rel. 34, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
E Gastrotropin (Gr) (Ileal lipid-binding protein) (ILBP).
N PABP6 OR ILBP
S Mus musculus (Mouse).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X NCBI_TaxID=10090;
N [1]
N SEQUENCE FROM N.A.
C STRAIN=DRA/2J; TISSUE=Liver;
C MEDLINE=94375529; Pubmed=8089185;
X Crossman M.W., Hault S.M., Gordon J.I.;
A "The mouse ileal lipid-binding protein gene: a model for studying
T axial patterning during gut morphogenesis.";
J. Cell Biol. 136:1547-1564(1994).
L - FUNCTION: ILEAL PROTEIN WHICH STIMULATES GASTRIC ACID AND
P PEPINOGEN SECRETION. SEEMS TO BE ABLE TO BIND TO BILE SALTS AND
BILIURUBINS.
C - SUBCELLULAR LOCATION: Cytoplasmic.
C - SIMILARITY: Belongs to the fatty-acid binding protein (FABP)
family.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sb-sib.ch).
EMBL; U00938; AAC27352.1; --
PIR; A54797; A54797.
HSSP; P10289; 1EAL.
MGI; 96565; Fabbp6.
InterPro; IPR000463; Fatty acid BP.
InterPro; IPR000566; Lipocalin_cyFABP.
Pfam; PF00061; lipocalin; 1
PRINIS; PR00178; FATTYACIDBP.
PROSITE; PS00214; FABP; 1.
Transport; Lipid-binding; Acetylation.
INIT_MET 0
BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT MOD_RES 1 1
SQ SEQUENCE 127 AA; 14355 MW; 0690BDD0A9CD922 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
b 91 VAEF 94

RESULT 16
SPEH_THEME STANDARD; PRT; 130 AA.
AC Q9WC3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE S-adenosylmethionine decarboxylase proenzyme (SC 4.1.1.50) (AdoMetDC)
DE (samDC) [Contains: S-adenosylmethionine decarboxylase beta chain; S-
DE adenosylmethionine decarboxylase alpha chain].
GN SPEH OR TM0655;
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=MSB / DSM 3109 / ATCC 43589;
RX MEDLINE=99387316; Pubmed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickley E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC - FUNCTION: Decarboxylation of S-adenosylmethionine provides the
CC aminopropyl moiety required for spermidine biosynthesis from
CC putrescine (By similarity).
CC - CATALYTIC ACTIVITY: S-adenosyl-L-methionine = (S)-deoxy-5-
CC adenosyl(3-aminopropyl) methylsulfonium salt + CO(2).
CC - COFACTOR: Pyruvoyl group (By similarity).
CC - SIMILARITY: Belongs to the prokaryotic AdoMetDC family. Subfamily
CC 1.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sb-sib.ch).
EMBL; AE001739; AAD35739.1; --
PIR; D72348; D72348.
TIGR; TM0655; --
HAWAP; MF_00464; --; 1.
InterPro; IPR003826; SAMDC.
Pfam; PF02675; AdoMetDC; 1.
KW Spermidine biosynthesis; Lyase; Decarboxylase; Zymogen; Pyruvate;
KW Complete proteome.
FT CHAIN 1 62 S-ADENOSYLMETHIONINE DECARBOXYLASE BETA
FT CHAIN 63 130 CHAIN (BY SIMILARITY)
FT CHAIN 63 130 S-ADENOSYLMETHIONINE DECARBOXYLASE ALPHA
FT SITE 62 63 CHAIN (BY SIMILARITY).
FT MOD_RES 63 63 CLEAVAGE (NONHYDROLYTIC) (BY SIMILARITY)
FT MOD_RES 63 63 CONVERTED TO A PYRUVOYL GROUP (BY
FT MOD_RES 63 63 SIMILARITY).
SQ SEQUENCE 130 AA; 14785 MW; 7659FE20A2019928 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 130;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QV 2 VAEF 5
Db 9 VAEF 12

RESULT 17
Y194_PYRAB STANDARD; PRT; 131 AA.
ID Y194_PYRAB

```

Q9V280;  
 16-OCT-2001 (Rel. 40, Created)  
 16-OCT-2001 (Rel. 40, Last sequence update)  
 10-OCT-2003 (Rel. 42, Last annotation update)  
 Hypothetical UPF0146 protein PYRAB01940.  
 PYRAB01940 OR PAB2224.  
 Pyrococcus abyssi.  
 Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 Pyrococcus.  
 NCBI\_TaxID=29292;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=GES / Orsay.  
 MEDLINE=29511545; PubMed=12622808;  
 Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,  
 Poch O., Prieur D., Querrelou J., Ripp R., Thierry J.-C.,  
 Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;  
 "An integrated analysis of the genome of the hyperthermophilic  
 archaeon Pyrococcus abyssi."  
 Mol. Microbiol. 47:1495-1512(2003).  
 -!- SIMILARITY: Belongs to the UPF0146 family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; AJ248283; CAB49118.1; -.  
 PIR; G75208; G75208.  
 HAMAP; MF\_00341; -; 1.  
 InterPro; IPR005353; UPF0146.  
 Pfam; PF03686; UPF0146; 1.  
 PIRSF; PIRSF016725; UCP016725; 1.  
 ProDom; PD021130; UPF0146; 1.  
 Hypothetical protein; Complete proteome.  
 SEQUENCE 131 AA; 14629 MW; D13F378187F832C3 CRC64;  
 -----  
 Query Match 95.0%; Score 19; DB 1; Length 131;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 / 2 VAEF 5  
 4 VAEF 7  
 -----  
 RESULT 18  
 ISB HALN1  
 RIBS HALN1 STANDARD; PRT; 133 AA.  
 Q9HSM5,  
 16-OCT-2001 (Rel. 40, Created)  
 16-OCT-2001 (Rel. 40, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)  
 (lumazine synthase) (Riboflavin synthase beta chain).  
 RIBB OR RIBB OR VNG0630G.  
 Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).  
 Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 Halobacteriaceae; Halobacterium.  
 NCBI\_TaxID=64091;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=20504493; PubMed=11016950;  
 Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
 Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sirogna J.,  
 Swartzell S., Weir D., Hall J.A., Dahl T.A., Welti R., Goo Y.A.,  
 Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 Maddocks D.G., Jablonki P.E., Krebs M.P., Angevine C.M., Dale H.,  
 Isenbarger T.A., Peck R.F., Polschroder M., Spidich J.L., Jung K.-H.,  
 Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,

RA Ebbardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
 RT "Genome sequence of Halobacterium Species NRC-1."  
 Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC -!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex  
 catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-  
 ribityl-amino-2,4(1H,3H)-pyrimidinedione and L-3,4-dihydroxy-2-  
 butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit  
 catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-  
 2,4(1H,3H)-pyrimidinedione with L-3,4-dihydroxy-2-butanone-4-  
 phosphate yielding 6,7-dimethyl-8-lumazine (by similarity).  
 CC CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1-D-ribityl)lumazine =  
 CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.  
 CC -!- PATHWAY: Riboflavin biosynthesis; last step.  
 CC -!- SIMILARITY: Belongs to the DMRL synthase family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; AE005011; AAG19133.1; -.  
 PIR; A84221; A84221.  
 DR HSSP; O66529; LHQK.  
 DR HAMAP; MF\_00178; -; 1.  
 DR InterPro; IPR002180; DMRL synthase.  
 DR Pfam; PF00885; DMRL synthase; 1.  
 DR ProDom; PD003664; DMRL synthase; 1.  
 KW Riboflavin biosynthesis; Transferase; Complete proteome.  
 SQ SEQUENCE 133 AA; 13719 MW; B9724DC24660D6CD CRC64;  
 -----  
 Query Match 95.0%; Score 19; DB 1; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 8 VAEF 11  
 -----  
 RESULT 19  
 YBGC ECOLI  
 ID YBGC ECOLI STANDARD; PRT; 134 AA.  
 AC P08959;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protein YBGC.  
 GN YBGC OR B0736 OR C0815 OR Z0904 OR ECS0771 OR SF0561 OR S0574.  
 OS Escherichia coli,  
 OS Escherichia coli O6,  
 OS Escherichia coli O157:H7, and  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia  
 OX NCBI\_TaxID=562, 217992, 83334, 623;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RF SPECIES=E.coli;  
 RC MEDLINE=87222192; PubMed=3294803;  
 RA Sun T.P., Webster R.E.;  
 RT "Nucleotide sequence of a gene cluster involved in entry of E colicine  
 and single-stranded DNA of infecting filamentous bacteriophages into  
 Escherichia coli."  
 RL J. Bacteriol. 169:2667-2674(1987).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP SPECIES=E.coli; STRAIN=K12 / MG1655;  
 RC MEDLINE=97426617; PubMed=9278503;  
 RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 MAU B., Shao Y.,  
 "The complete genome sequence of *Escherichia coli* K-12.",  
 Science 277:1453-1474(1997).  
 [3]  
 SEQUENCE FROM N.A.  
 SPECIES=E.coli; STRAIN=K12;  
 MEDLINE=97061202; PubMed=8905232;  
 Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 Sampei G., Seki I., Tagami H., Takemoto K., Mada C., Yamamoto Y.,  
 Yano M., Horiuchi T.;  
 "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome  
 corresponding to the 12.7-28.0 min region on the linkage map.";  
 DNA Res. 3:137-155(1996).  
 [4]  
 SEQUENCE FROM N.A.  
 SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 MEDLINE=22388234; PubMed=12471157;  
 Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic *Escherichia coli*.";  
 Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 [5]  
 SEQUENCE FROM N.A.  
 SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 MEDLINE=21074935; PubMed=11206553;  
 Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 Welch R.A., Blattner F.R.;  
 "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";  
 Nature 409:523-533(2001).  
 [6]  
 SEQUENCE FROM N.A.  
 SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;  
 MEDLINE=21156231; PubMed=11258796;  
 Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,  
 Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
 O157:H7 and genomic comparison with a laboratory strain K-12.";  
 DNA Res. 8:11-22(2001).  
 [7]  
 SEQUENCE OF 1-20 FROM N.A.  
 SPECIES=E.coli; STRAIN=K12 / MG1655;  
 Kim K., Allen E., Araujo R., Aparicio A.M., Botstein D.,  
 Cherry M., Chung E., Dietrich F., Duncan M., Federspiel N.,  
 Kalman S., Komp C., Lashkari D., Lew H., Lin D., Namath A.,  
 Oefner P., Davis R.;  
 Submitted (JUL-1995) to the ENBL/GenBank/DBJ databases.  
 [8]  
 SEQUENCE FROM N.A.  
 SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
 MEDLINE=22272406; PubMed=12384590;  
 Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
 Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 Yu J.;  
 "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity  
 through comparison with genomes of *Escherichia coli* K12 and O157.";  
 Nucleic Acids Res. 30:4432-4441(2002).  
 [9]  
 SEQUENCE FROM N.A.  
 SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 MEDLINE=22590274; PubMed=12704152;  
 Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 Schwartz D.C., Blattner F.R.;  
 "Complete genome sequence and comparative genomics of *Shigella*  
 flexneri serotype 2a strain 2457T.";  
 Infect. Immun. 71:2775-2786(2003).  
 [10]  
 IDENTIFICATION BY MASS SPECTROMETRY.  
 SPECIES=E.coli;  
 MEDLINE=99420866; PubMed=10493123;  
 Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;  
 "Enrichment of low abundance proteins of *Escherichia coli* by  
 hydroxyapatite chromatography.";  
 Electrophoresis 20:2181-2195(1999).  
 CC -1- SIMILARITY: BELONGS TO THE 4-HYDROXYBENZYOYL-COA THIOESTERASE  
 FAMILY. STRONG, TO H-INFLUENZAE HI0396.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 EMBL; M16489; AAA83918.1; -;  
 EMBL; AE000177; AAC73830.1; -;  
 EMBL; D90713; BAA35402.1; -;  
 EMBL; AE016757; AAN79288.1; -;  
 EMBL; AE005252; AAG55072.1; -;  
 EMBL; AP002553; BAB34194.1; -;  
 EMBL; U30934; AAA74398.1; -;  
 EMBL; AE015086; AAN42205.1; -;  
 EMBL; AE016979; AAP16078.1; -;  
 PIR; A25980; WMEC15.  
 PIR; C90725; C90725.  
 PIR; D85576; D85576.  
 ECoGene; EG11110; YbgC.  
 InterPro; IPR008272; 4HBCOA\_thioest\_AS.  
 InterPro; IPR008684; 4HBCOA\_thioestse.  
 InterPro; IPR006683; Thioest\_suppl.  
 Pfam; PF03061; 4HBT; 1.  
 TIGRFAMs; TIGR00051; TIGR00051; 1.  
 PROSITE; PS01328; 4HBCOA\_THIOESTERASE; 1.  
 KW Hydrolyase; Complete proteome.  
 FT ACT SITE 18 BY SIMILARITY.  
 SQ SEQUENCE 134 AA; 15562 MW; C4582B6EC3BE989 CRC64;  
 Query Match 95.0%; Score 19; DB 1; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 DB 129 VAEF 132  
 RESULT 20  
 RISB METH  
 ID RISB METH STANDARD; PRT; 139 AA.  
 AC C27443;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)  
 DE (Lumazine synthase) (Riboflavin synthase beta chain).  
 GN RIBH OR WHI1390.  
 OS Methanobacterium thermoautotrophicum.  
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
 OC Methanobacteriaceae; Methanothermobacter.

Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D., Richardson D.L., Kerlavage A.R., Graham D.E., Kyripides N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.P., McDonald L., Utterback T., Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M., Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus";  
Nature 390:364-370(1997).  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
-----  
EMBL; AF001064; AAB90658.1; -.  
PIR; B69323; B69323.  
TIGR; AF0586; -.  
KW Hypothetical protein; Complete proteome.  
DR HYPOTHELTICAL PROTEIN; Complete proteome.  
CC SEQUENCE 140 AA; 16175 MW; D3BFE3A62A05901 CRC64;  
-----  
Query Match 95.0%; Score 19; DB 1; Length 140;  
Best Local Similarity 100.0%; Pred.No. 2.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 VAEF 5  
|||||  
DB 65 VAEF 68  
|||

RESULT 22  
RISB ARCFCU  
ID RISB ARCFCU STANDARD; PRT; 143 AA.  
AC Q28152;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase) (lumazine synthase) (Riboflavin synthase beta chain).  
GN RIBH OR RIBE OR AF2128.  
OS Archaeoglobus fulgidus.  
OC Archaea, Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
[1]  
SEQUENCE FROM N.A.  
RP STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Kleck H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D., Richardson D.L., Kerlavage A.R., Graham D.E., Kyripides N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.P., McDonald L., Utterback T., Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M., Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus";  
Nature 390:364-370(1997).  
-----  
FUNCTION: Riboflavin synthase is a bifunctional enzyme complex catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-ribityl-amino-2,4'(IH,3H)-pyrimidinethione and L-3,4-dihydroxy-2-butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit

CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-  
 CC 2,4(H,3H)-pyrimidinone with L-3,4-dihydroxy-2-butanone-4-  
 CC phosphate yielding 6,7-dimethyl-8-lumazine (By similarity).  
 CC -!- CATALYTIC ACTIVITY: 2,6,7-dimethyl-8-(1'-D-ribityl)lumazine =  
 CC riboflavin + 4-(1'-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.  
 CC -!- PATHWAY: Riboflavin biosynthesis; last step.  
 CC -!- SIMILARITY: Belongs to the DMRL synthase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

CC EMBL; AE000957; AAB89124.1; -;  
 CC PIR; H69515; H69515.  
 CC HSSP; P11998; 1RVV.  
 CC TIGR; AF2128; -;  
 CC HAVAP; MF\_00178; -; 1.  
 CC InterPro; IPR002180; DMRL synthase.  
 CC Pfam; PF00885; DMRL synthase; 1.  
 CC ProDom; PD003664; DMRL synthase; 1.  
 CC TIGRFAMs; TIGR00114; ribH; 1  
 CC Riboflavin biosynthesis; Transferase; Complete proteome.  
 CC Riboflavin biosynthesis; Transferase; Complete proteome.  
 CC SEQUENCE 143 AA; 15645 MW; BBA3141206905410 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 10 VAEF 13

## RESULT 23

CC RL13 HALMA STANDARD; PRT; 145 AA.  
 CC  
 CC P29198;  
 CC 01-DEC-1992 (Rel. 24, Created)  
 CC 01-DEC-1992 (Rel. 24, Last sequence update)  
 CC 15-MAR-2004 (Rel. 43, Last annotation update)  
 CC 50S ribosomal protein L13P (Hmal13).  
 CC RL13P.  
 CC Halocarcula marismortui (Halobacterium marismortui).  
 CC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 CC Halobacteriaceae; Halocaula.  
 CC NCBI\_TaxID=2238;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE=92105119; PubMed=1840597;  
 CC Kroeber W.J., Arndt E.;  
 CC "Halobacterial S9 operon. Three ribosomal protein genes are  
 CC cotranscribed with genes encoding a tRNA(Leu), the enolase, and a  
 CC putative membrane protein in the archaeobacterium Halocaula  
 CC (Halobacterium) marismortui.";  
 CC J. Biol. Chem. 266:24573-24579 (1991).  
 CC [2]  
 CC X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 CC STRAIN=ATCC 43049;  
 CC MEDLINE=20396344; PubMed=10937989;  
 CC Ban N., Nissen P., Hansen J., Moore P.B., Steltz T.A.;  
 CC "The complete atomic structure of the large ribosomal subunit at 2.4  
 CC A resolution.";  
 CC Science 289:905-920 (2000).  
 CC -!- SIMILARITY: Belongs to the L13P family of ribosomal proteins.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

CC EMBL; M76567; AAA73097.1; -;  
 CC PIR; B41715; B41715.  
 CC PDB; 1FFK; 26-SEP-01.  
 CC PDB; 1K8A; 19-JUL-02.  
 CC PDB; 1K9N; 19-JUL-02.  
 CC PDB; 1KD1; 19-JUL-02.  
 CC PDB; 1M1K; 23-AUG-02.  
 CC PDB; 1M90; 06-SEP-02.  
 CC InterPro; IPR005822; Ribosomal L13.  
 CC InterPro; IPR005755; Ribosomal L13e/a.  
 CC Pfam; PF00572; Ribosomal L13; 1.  
 CC TIGRFAMs; TIGR01077; L13 A\_E; 1.  
 CC PROSITE; PS00783; RIBOSOMAL L13; 1.  
 CC Ribosomal protein; 3D-structure.  
 CC SEQUENCE 145 AA; 16228 MW; 069CE666662AE3BC CRC64;

Query Match 95.0%; Score 19; DB 1; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 3 VAEF 6

## RESULT 24

CC CLM4 MOUSE STANDARD; PRT; 148 AA.  
 CC  
 CC QJME3; QJCE31; Q9D1B3;  
 CC 16-OCT-2001 (Rel. 40, Created)  
 CC 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Calmodulin 4 (Calcium-binding protein Dd112).  
 CC CLM4 OR Dd112.  
 CC Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 CC NCBI\_TaxID=10090;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC Ikegawa S., Nakamura Y.;  
 CC "Dd112, a novel mouse gene implicated in the early stage of ectopic  
 CC ossification.";  
 CC Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=C57BL/6J; TISSUE=Tongue;  
 CC MEDLINE=21085660; PubMed=11217851;  
 CC Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 CC Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 CC Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakawa I.,  
 CC Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 CC Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 CC Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 CC Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 CC Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 CC Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 CC Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 CC Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 CC Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 CC Lyons P., Marchionni L., Maehama J., Mazzarelli J., Mombaerts P.,  
 CC Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 CC Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 CC Suzuki H., Toyono-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 CC Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 CC Hayashizaki Y.;  
 CC "Functional annotation of a full-length mouse cDNA collection.";  
 CC Nature 409:685-690 (2001).  
 CC -!- FUNCTION: Implicated in the early stage of ectopic ossification.



JB Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).  
 JS Prionace glauca (Blue shark).  
 XC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
 XC Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes;  
 XC Carcharhinidae; Prionace.  
 X NCBI\_TaxID=7815;  
 IN [1]  
 P SEQUENCE.  
 X MEDLINE=8929032; PubMed=2500367;  
 A Calabrese L., Politicelli F., O'Neill P., Galtieri A., Barra D.,  
 A Schinina M.E., Bossa F.;  
 T "Substitution of arginine for lysine 134 alters electrostatic  
 T parameters of the active site in shark Cu,Zn superoxide dismutase.";  
 L FEBS Lett. 250:49-52(1989).  
 C -1- FUNCTION: Destroys radicals which are normally produced within the  
 C cells and which are toxic to biological systems.  
 C -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
 C -1- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit.  
 C -1- SUBUNIT: Homodimer.  
 C -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 C -1- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.  
 R PIR: S04623; S04623.  
 R HSP; P00442; ICBJ.  
 R InterPro; IPR001424; SOD\_CU\_ZN.  
 R Pfam; PF00080; sodcu; 1.  
 R PRINTS; PR00068; CUZNDISMTASE.  
 R PRODOM; PD000469; SOD\_CU\_ZN; 1.  
 R PROSITE; PS00087; SOD\_CU\_ZN\_1; 1.  
 R PROSITE; PS00332; SOD\_CU\_ZN\_2; 1.  
 W Antioxidant; Oxidoreductase; Metal-binding; Copper; Zinc.  
 T METAL 44 44  
 T METAL 46 46 COPPER.  
 T METAL 61 61 COPPER AND ZINC.  
 T METAL 69 69 ZINC.  
 T METAL 78 78 ZINC.  
 T METAL 81 81 ZINC.  
 T METAL 118 118 COPPER.  
 T METAL 118 118  
 T DISULFID 55 144 BY SIMILARITY.  
 T SEQUENCE 152 AA; 15840 MW; 6617642A4F23CSAE CRC64;  
 Q  
 Query Match 95.0%; Score 19; DB 1; Length 152;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 2 VAEF 5  
 b 92 VAEF 95  
 RESULT 28  
 XLS\_SULSO  
 ID RISS\_SULSO STANDARD; PRT; 154 AA.  
 AC Q980B5;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)  
 DE (lumazine synthase) (Riboflavin synthase beta chain).  
 RN RIB OR SSO0400.  
 JS Sulfolobus solfataricus.  
 XC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 XC Sulfolobus.  
 X NCBI\_TaxID=2287;  
 IN [1]  
 P SEQUENCE FROM N.A.  
 X STRAIN=ATCC 35092 / DSM 1617 / P2;  
 X MEDLINE=21332296; PubMed=11427726;  
 X She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 X Awaye M.J., Chan-Weifer C.-Y., Clausen I.G., Curtis B.A.,  
 X De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,  
 X Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 X Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 X Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,

RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;  
 R "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 CC -1- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex  
 CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-  
 CC ribityl-amino-2,4(1H,3H)-pyrimidinedione and L-3,4-dihydroxy-2-  
 CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit  
 CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-  
 CC amino-2,4(1H,3H)-pyrimidinedione with L-3,4-dihydroxy-2-  
 CC butanone-4-phosphate yielding 6,7-dimethyl-8-lumazine (By  
 CC similarity).  
 CC -1- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1-D-ribityl)lumazine =  
 CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.  
 CC -1- PATHWAY: Riboflavin biosynthesis; last step.  
 CC -1- SIMILARITY: Belongs to the DMRL synthase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE006673; AAK40729.1; --  
 CC PIR; B90184; B90184.  
 CC HAMAP; MF\_00178; --; 1.  
 CC InterPro; IPR002180; DMRL synthase.  
 CC Pfam; PF00885; DMRL synthase; 1.  
 CC PRODOM; PD003664; DMRL synthase; 1.  
 CC TRIGRAMS; TRIGR00114; ribh; 1.  
 KW Riboflavin biosynthesis; Transferase; Complete proteome.  
 SQ SEQUENCE 154 AA; 17247 MW; ADF956A2C723210 CRC64;  
 Query Match 95.0%; Score 19; DB 1; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 Db 12 VAEF 15  
 RESULT 29  
 PCP\_HASIN  
 ID PCP\_HASIN STANDARD; PRT; 155 AA.  
 AC P10325;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Outer membrane lipoprotein PCP precursor (15 kDa lipoprotein) (PAL  
 DE cross-reacting lipoprotein).  
 GN PCP OR LPP OR H1579.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 CX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88115138; PubMed=2828309;  
 RA Deich R.A., Metcalf B.J., Finn C.W., Farley J.E., Green B.A.;  
 RT Cloning of genes encoding a 15,000-dalton peptidoglycan-associated  
 RT outer membrane lipoprotein and an antigenically related 15,000-dalton  
 RT protein from Haemophilus influenzae.";  
 RL J. Bacteriol. 170:489-498(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Karlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

```
1 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
2 Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
3 Fine L.D., Fritchman J.L., Fuhrmann J.L., Geognagen N.S.M.,
4 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
5 Venter J.C.;
6 "Whole-genome random sequencing and assembly of Haemophilus influenzae
7 Rd.";
8 Science 269:496-512(1995).
9 -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
10 anchor.
11 -!- SIMILARITY: TO E-COLI AND S-TYPHIMURIUM SLYB AND TO
12 Y-ENTEROCOLITICA PCP.
13
14 This SWISS-PROT entry is copyright. It is produced through a collaboration
15 between the Swiss Institute of Bioinformatics and the EMBL outstation -
16 the European Bioinformatics Institute. There are no restrictions on its
17 use by non-profit institutions as long as its content is in no way
18 modified and this statement is not removed. Usage by and for commercial
19 entities requires a license agreement (See http://www.isb-sib.ch/announce/
20 or send an email to license@isb-sib.ch).
```

```
EMBL; M18877; AA224938.1; -
EMBL; U32832; AAC23228.1; -
PIR; I64130; I64130.
TIGR; H11579; -
InterPro; IPR000437; Prok lipoprot S.
PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
Outer membrane; Lipoprotein; Signal; Complete proteome; Palmitate.
SIGNAL 1 18
CHAIN 19 155
OUTER MEMBRANE LIPOPROTEIN PCP.
LIPID 19 19
N-palmitoyl cysteine.
LIPID 19 19
S-diacylglycerol cysteine.
CONFLICT 135 143 CSLVAEFVF -> VAGRRVRI (IN REF. 1).
SEQUENCE 155 AA; 15425 MW; D7880327FCFC985 CRC64;
```

Query Match 95.0%; Score 19; DB 1; Length 155;  
Best Local Similarity 100.0%; Pred. NO. 2.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
138 VAEF 141

```
RESULT 30
AP TAROF
D RAP TAROF STANDARD; PRT; 157 AA.
C O49065;
T 30-MAY-2000 (Rel. 39, Created)
I 30-MAY-2000 (Rel. 39, Last sequence update)
I 15-MAR-2004 (Rel. 43, Last annotation update)
S Root allergen protein (RAP).
E Taraxacum officinale (Common dandelion).
S Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
S Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
S campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae;
S Taraxacum.
X NCBI_TAXID=50225;
P SEQUENCE FROM N.A.
N TISSUE=Root.
A Xu X.-Y., Bewley J.D., Greenwood J.S.;
L Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
C -!- ALLERGEN: Causes an allergic reaction in human.
C -!- SIMILARITY: Belongs to the BetV family.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
```

```
CC EMBL; AF036931; AAB92255.1; -
DR HSSP; O24248; IR09.
DR InterPro; IPR000916; Bet_v_I.
DR Pfam; PF00407; Bet_v_I_1.
DR PRINTS; PRO0634; BETALLERGEN.
DR PROSITE; PS00451; PATHOGENESIS BETVI; 1.
KW Allergen; Plant defense; Pathogenesis-related protein.
SQ SEQUENCE 157 AA; 17040 MW; 5892AB85938A7E0 CRC64;
```

Query Match 95.0%; Score 19; DB 1; Length 157;  
Best Local Similarity 100.0%; Pred. NO. 2.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
3 VAEF 6

```
RESULT 31
RISB SULTO
ID RISB SULTO STANDARD; PRT; 157 AA.
AC Q975M5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)
DE (lumazine synthase) (Riboflavin synthase beta chain).
GN RIBH OR ST0394.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=2145156; PubMed=11572479;
RA Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaei A., Kosugi H., Rosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
CC -!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
CC ribityl-amino-2,4(1H,3H)-pyrimidinone and L-3,4-dihydroxy-2-
CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-
CC amino-2,4(1H,3H)-pyrimidinone with L-3,4-dihydroxy-2-
CC butanone-4-phosphate yielding 6,7-dimethyl-8-lumazine (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1-D-ribityl) lumazine =
CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
CC -!- PATHWAY: Riboflavin biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the DMRL synthase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

EMBL; AF000982; BAB65375.1; -  
HMAP; MF 00478; -; 1.  
InterPro; IPR002180; DMRL synthase.  
Pfam; PF00885; DMRL synthase; 1.  
ProDom; PD003664; DMRL synthase; 1.  
TIGRFAMs; TIGR00114; ribH; 1.

W Riboflavin biosynthesis; Transferase; Complete proteome.  
 Q SEQUENCE 157 AA; 17501 MW; 529558077EL64A18 CRC64;  
 Query Match 95.0%; Score 19; DB 1; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 2 VAEF 5  
 ||||  
 b 14 VAEF 17  
 ||||  
 RESULT 32  
 IGSS DROME STANDARD; PRT; 163 AA.  
 C P07701; QSVET5;  
 T 01-APR-1988 (Rel. 07, Created)  
 T 01-APR-1988 (Rel. 07, Last sequence update)  
 T 10-OCT-2003 (Rel. 42, Last annotation update)  
 E Salivary glue protein sgs-5 precursor.  
 N SG55 OR CG7596.  
 S Drosophila melanogaster (Fruit fly).  
 C Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 C Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 C Ephydroidea; Drosophilidae; Drosophila.  
 X NCBI\_TaxID=7227;  
 [1]  
 P SEQUENCE FROM N.A.  
 C STRAIN=Oregon-R;  
 X MEDLINE=87086754; PubMed=3098981;  
 X Shore E.M., Guild G.M.;  
 T "Larval salivary gland secretion proteins in Drosophila structural  
 analysis of the sgs-5 gene.";  
 J. Mol. Biol. 190:149-158 (1986).  
 [2]  
 P SEQUENCE FROM N.A.  
 C STRAIN=Berkley;  
 X MEDLINE=20196006; PubMed=10731132;  
 X Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 A Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 A George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 A Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 A Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 A Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 A Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 A Ballew R.M., Basu A., Baxendale J., Bayraktarglu L., Beasley E.M.,  
 A Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 A Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 A Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 A Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 A de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 A Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 A Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 A Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 A Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 A Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 A Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 A Jafarizadeh M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 A Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Klupp D., Lai Z.,  
 A Laake P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 A Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 A Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 A Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 A Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paolel J.M.,  
 A Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 A Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,  
 A Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,  
 A Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 A Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 A Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 A Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 A Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 A Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RL "The genome sequence of Drosophila melanogaster.";  
 CC Science 287:2185-2195 (2000).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X04269; CAA27820.1; -;  
 DR EMBL; AE003718; AAF55436.1; -;  
 DR PIR; A24504; A24504.  
 DR FlyBase; FBgn0003375; Sgs5.  
 KW Signal.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 163 SALIVARY GLUE PROTEIN SGS-5.  
 SQ SEQUENCE 163 AA; 18821 MW; 3A476FD3B06D864 CRC64;  
 Query Match 95.0%; Score 19; DB 1; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 ||||  
 DB 114 VAEF 117  
 ||||  
 RESULT 33  
 UTR5\_YEAST STANDARD; PRT; 166 AA.  
 AC P32630;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE UTR5 protein (Unknown transcript 5 protein)  
 GN UTR5 OR YEL035C OR SYGP-ORE27.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CX NCBI\_TaxID=4932;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=B-6441;  
 RX MEDLINE=94016558; PubMed=8411151;  
 RA Melnick L., Sherman F.;  
 RT "The gene clusters ARC and COR on chromosomes 5 and 10, respectively,  
 RL of Saccharomyces cerevisiae share a common ancestry.";  
 J. Mol. Biol. 233:372-388 (1993).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=S288C / AB972;  
 RX MEDLINE=97313264; PubMed=9169868;  
 RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,  
 RA Araujo R., Aviles E., Berno A., Guzman E., Hartzell G.,  
 RA Cherry J.M., Chung E., Duncan M., Kayser A., Komp C., Lashkari D., Lew H.,  
 RA Hunkle-Smith S., Hyman R.W., Kayaer A., Namath A., Norgren R., Oefner P.,  
 RA Lin D., Mosedale D., Nakahara K., Sahl P., Schramm S., Shogren T.,  
 RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,  
 RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";  
 Nature 387:78-81 (1997).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

EMBL; L22173; AAA34936.1; -  
 EMBL; S65964; AAD13970.1; -  
 EMBL; S66120; AAB28442.1; -  
 EMBL; U18779; AAB65007.1; -  
 PIR; S50509; S50509.  
 GenBank; U18779; S50509.  
 SGD; S0000761; UTR5.  
 CONFLICT 1 20  
 MSRYGKNLVHYIIVHDDQR -> MRDSNVK-SVFFCALYN  
 RGNTTIN (IN REF. 1).

SEQUENCE 166 AA; 19335 MW; A58EB8D0FAGD3DEA CRC64;

Query Match 95.0%; Score 19; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
 ||||  
 92 VAEF 95

## SUIT 34

B\_RHOBA STANDARD; PRT; 169 AA.  
 P59332; OTUKV3;  
 15-MAR-2004 (Rel. 43, Created)  
 15-MAR-2004 (Rel. 43, Last sequence update)  
 15-MAR-2004 (Rel. 43, Last annotation update)  
 Single-strand binding protein (SSB) (Helix-destabilizing protein).  
 SSB OR RB9917.  
 Rhodopirellula baltica.  
 Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
 Planctomycetaceae; Pirellula.  
 NCBI\_TaxID=117;  
 [1]

SEQUENCE FROM N.A.

STRAIN=1;  
 MEDLINE=22735913; PubMed=12835416;  
 Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
 Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,  
 Schlesner H., Amann R., Reinhardt R.;  
 "Complete genome sequence of the marine planctomycete Pirellula sp.  
 strain 1.";  
 Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).

-!- FUNCTION: This protein is essential for replication of the  
 chromosome. It is also involved in DNA recombination and repair  
 (By similarity).

-!- SIMILARITY: Contains 1 SSB domain.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; BX294150; CAD76529.1; -  
 PROSITE; PS50935; SSB; 1.

DNA-binding; DNA repair; DNA replication; Complete proteome.

DOMAIN 4 107 SSB.

DOMAIN 113 121 POLY-GLY.

SEQUENCE 169 AA; 18108 MW; AABF93E0FAP51F287 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 169;

Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
 ||||  
 28 VAEF 31

## RESULT 35

BFL1 MOUSE STANDARD; PRT; 172 AA.  
 ID BFL1 MOUSE  
 AC Q07440;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Bcl-2-related protein A1 (BFL-1 protein) (Hemopoietic-specific early  
 response protein) (A1-A).  
 DN BCL2A1 OR BCL2A1A OR BFL1 OR A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 SEQUENCE FROM N.A.  
 STRAIN=CBA/J; TISSUE=Bone marrow;  
 MEDLINE=93346743; PubMed=8345191;  
 RA Lin E.Y., Orlofsky A., Berger M.S., Prystowsky M.B.;  
 "Characterization of A1, a novel hemopoietic-specific early-response  
 gene with sequence similarity to bcl-2.";  
 J. Immunol. 151:1979-1988 (1993).  
 [2]

SEQUENCE FROM N.A.

STRAIN=129/SV; TISSUE=Liver;  
 MEDLINE=98307518; PubMed=9645611;  
 RA Hatakeyama S., Hamasaki A., Negishi I., Loh D.Y., Sendo F.,  
 Nakayama K., Nakayama K.-I.;  
 "Multiple gene duplication and expression of mouse bcl-2-related  
 genes, A1.";  
 Int. Immunol. 10:631-637 (1998).  
 CC -!- FUNCTION: Retards apoptosis induced by IL-3 deprivation. May  
 function in the response of hemopoietic cells to external signals  
 and in maintaining endothelial survival during infection.  
 CC -!- SUBCELLULAR LOCATION: Intracellular.  
 CC -!- TISSUE SPECIFICITY: Expressed in hemopoietic tissues, including  
 bone marrow, spleen and thymus.  
 CC -!- INDUCTION: By granulocyte-macrophage colony-stimulating factor  
 and LPS in macrophages.  
 CC -!- SIMILARITY: Contains 1 Bcl-2 homology 1 (BH1) domain.  
 CC -!- SIMILARITY: Contains 1 Bcl-2 homology 2 (BH2) domain.  
 CC -!- SIMILARITY: Belongs to the Bcl-2 family.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L16462; AAA16886.1; -

EMBL; U23774; AAB97953.1; -

EMBL; U23773; AAB97953.1; JOINED.

PIR; I49449; I49449.

HSP; Q07817; IMAZ.

MGB; MGI:102887; Bcl2ala.

InterPro; IPR000712; Bcl2\_BH.

InterPro; IPR002475; BCL2\_family.

Pfam; PF00452; Bcl-2; 1.

SMART; SM00337; BCL; 1.

PROSITE; PS50062; BCL2\_FAMILY; 1.

PROSITE; PS01080; BH1; 1.

PROSITE; PS01258; BH2; 1.

KW Apoptosis.

DOMAIN 24 33 ALA/PRO-RICH.

DOMAIN 77 97 BH1.

DOMAIN 132 147 BH2.

SEQUENCE 172 AA; 19914 MW; 37AD35818E756488 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 172;

Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
122 VAEF 125

RESULT 36  
FL1\_HUMAN STANDARD; PRT; 175 AA.

C Q16548; Q99524;  
T 01-NOV-1997 (Rel. 35, Last sequence update)  
T 15-MAR-2004 (Rel. 43, Last annotation update)  
E Bcl-2-related protein A1 (BFL-1 protein) (Hemopoietic-specific early  
E response protein) (GRS protein).  
N BCL2A1 OR BFL1 OR GRS OR BCL2L5.  
S Homo sapiens (Human).  
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
X NCBI\_TaxID=9606;  
[1]  
N SEQUENCE FROM N.A.  
P TISSUE=Umbilical vein;  
C MEDLINE=96184764; PubMed=8605321;  
A Karsan A., Yee E., Kaushansky K., Harlan J.M.;  
T "Cloning of human Bcl-2 homologue: inflammatory cytokines induce  
T human A1 in cultured endothelial cells.";  
L Blood 87:3089-3096(1996).  
[2]  
N SEQUENCE FROM N.A.  
P TISSUE=Liver;  
C MEDLINE=96088895; PubMed=7478596;  
X Choi S.S., Park I.-C., Yun J.W.; Sung Y.C., Hong S.-I., Shin H.-S.;  
T "A novel Bcl-2 related gene, Bfl-1, is overexpressed in stomach  
T cancer and preferentially expressed in bone marrow.";  
L Oncogene 11:1693-1698(1995).  
[3]  
N SEQUENCE FROM N.A.  
P TISSUE=I-cell;  
X Kenny J.J., Knobloch T.J., Augustus M., Carter K.C., Rosen C.A.,  
A Lang J.C.;  
T "GRS, a novel member of the Bcl-2 gene family, is highly expressed in  
T multiple cancer cell lines and in normal leukocytes.";  
L Oncogene 14:1997-1001(1997).  
[4]  
N SEQUENCE FROM N.A.  
P TISSUE=Skin;  
X MEDLINE=92388257; PubMed=12477932;  
A Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
A Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
A Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
A Hopkins R.F., Jordan H., Moore J., Max S.L., Wang J., Hsieh F.,  
A Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
A Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
A Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
A Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
A Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
A Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
A Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
A Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
A Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
A Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
A Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
A Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,  
A Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
T "Generation and initial analysis of more than 15,000 full-length  
T human and mouse cDNA sequences.";  
U Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
C -!- FUNCTION: Retards apoptosis induced by IL-3 deprivation. May  
C function in the response of hemopoietic cells to external signals  
C and in maintaining endothelial survival during infection (By  
C similarity).  
C -!- SUBCELLULAR LOCATION: Intracellular.

CC -!- TISSUE SPECIFICITY: Seems to be restricted to the hematopoietic  
CC compartment. Expressed in peripheral blood, spleen, and bone  
CC marrow, at moderate levels in lung, small intestine and testis, at  
CC a minimal levels in other tissues. Also found in vascular smooth  
CC muscle cells and hematopoietic malignancies.  
CC -!- INDUCTION: By phorbol ester and inflammatory cytokines, such as  
CC TNF-alpha, or IL-1-beta, but not by growth factors.  
CC -!- SIMILARITY: Contains 1 Bcl-2 homology 1 (BH1) domain.  
CC -!- SIMILARITY: Contains 1 Bcl-2 homology 2 (BH2) domain.  
CC -!- SIMILARITY: Belongs to the Bcl-2 family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC  
CC EMBL; U29680; AAC50438.1; -;  
CC EMBL; U27467; AAC50288.1; -;  
CC EMBL; Y09397; CAA70566.1; -;  
CC EMBL; BC016281; AAH16281.1; -;  
CC PIR; I39055; I39055.  
CC HSP; P33583; IAF3.  
CC Genew; HGNC:1991; BCL2A1.  
CC MIM; 601056; -;  
CC GO; GO:0008189; F:Apoptosis inhibitor activity; TAS.  
CC GO; GO:0006916; P:anti-apoptosis; TAS.  
CC InterPro; IPR000712; Bcl2\_BH.  
CC InterPro; IPR002475; BCL2\_family.  
CC Pfam; PF00452; Bcl-2; 1.  
CC SMART; SM00337; BCL; 1.  
CC PROSITE; PS0062; BCL2\_FAMILY; 1.  
CC PROSITE; PS01080; BH1; 1.  
CC PROSITE; PS01258; BH2; 1.  
CC KW Apoptosis.  
FT DOMAIN 24 33 ALA/PRO-RICH.  
FT DOMAIN 77 97 BH1.  
FT DOMAIN 132 147 BH2.  
FT CONFLICT 72 72 N -> T (IN REF. 3).  
FT CONFLICT 107 107 Q -> H (IN REF. 3).  
SQ SEQUENCE 175 AA; 20132 MW; 329D98AF2BE07A0D CRC64;  
Query Match 95.0%; Score 19; DB 1; Length 175;  
Best Local Similarity 100.0%; Pred.No. 3.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 VAEF 5  
Db 122 VAEF 125  
RESULT 37  
FR12\_RANCA STANDARD; PRT; 176 AA.  
ID -!- RANCA  
AC P07798;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ferritin, middle subunit (Ferritin M) (Ferritin X) (Ferritin H').  
OS Rana catesbeiana (Bull frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana.  
CX NCBI\_TaxID=8400;  
[1]  
N SEQUENCE FROM N.A.  
P MEDLINE=97222424; PubMed=3495534;  
X Dickey L.F., Sreedharan S., Theil E.C., Didsbury J.R., Wang Y.-H.,  
X Kaufman R.E.;  
T "Differences in the regulation of messenger RNA for housekeeping and  
T specialized-cell ferritin. A comparison of three distinct ferritin  
T complementary DNAs, the corresponding subunits, and identification of

the first processed in amphibia.";  
J. Biol. Chem. 262:7901-7907(1987).

[2]

X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).

MEDLINE=9367924; PubMed=10439069;

Ha Y., Shi D., Small G.W., Thell E.C., Allewell N.M.;

"Crystal structure of bullfrog M ferritin at 2.8 A resolution:

analysis of subunit interactions and the binuclear metal center.";

J. Biol. Inorg. Chem. 4:243-256(1999).

-!- FUNCTION: Ferritin is an intracellular molecule that stores iron

in a soluble, nontoxic, readily available form. The functional

molecule, which is composed of 24 chains, is roughly spherical and

contains a central cavity into which the polymeric ferric iron

core is deposited.

-!- MISCELLANEOUS: THERE ARE THREE TYPES OF FERRITIN SUBUNITS: L, M

AND H CHAINS IN AMPHIBIA.

-!- SIMILARITY: Belongs to the ferritin family.

-!- SIMILARITY: Contains 1 ferritin-like diiron domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; J02724; AAA49525.1; -.

PIR; C27805; C27805.

PDB; 1MER; 22-JUN-99.

InterPro; IPR001519; Ferritin.

InterPro; IPR008331; Ferritin\_Dps.

InterPro; IPR009040; Ferritin\_Like.

Pfam; PF00210; Ferritin; 1.

ProDom; PD000971; Ferritin; 1.

PROSITE; PS00540; FERRITIN\_1; 1.

PROSITE; PS00204; FERRITIN\_2; 1.

PROSITE; PS50905; FERRITIN\_LIKE; 1.

Iron storage; Iron; Metal-Binding; 3D-structure.

DOMAIN 7 156 FERRITIN-LIKE DIIRON.

METAL 24 24 IRON (BY SIMILARITY).

METAL 58 58 IRON (BY SIMILARITY).

METAL 59 59 IRON (BY SIMILARITY).

METAL 62 62 IRON (BY SIMILARITY).

METAL 104 104 IRON (BY SIMILARITY).

METAL 138 138 IRON (BY SIMILARITY).

HELIX 11 39

TURN 41 43

TURN 74 74

STRAND 82 82

HELIX 93 120

TURN 121 122

TURN 124 133

TURN 134 134

TURN 135 154

TURN 155 159

TURN 161 170

TURN 171 171

SEQUENCE 176 AA; 20592 MW; A9F0F5BEB8584D46 CRC64;

-----

Query Match 95.0%; Score 19; DB 1; Length 176;

Best Local Similarity 100.0%; Pred. No. 3.1e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

-----

2 VAEF 5

48 VAEF 51

-----

RESULT 38

10\_THETN STANDARD; PRT; 177 AA.

-----

the first processed in amphibia.";

J. Biol. Chem. 262:7901-7907(1987).

[2]

X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).

MEDLINE=9367924; PubMed=10439069;

Ha Y., Shi D., Small G.W., Thell E.C., Allewell N.M.;

"Crystal structure of bullfrog M ferritin at 2.8 A resolution:

analysis of subunit interactions and the binuclear metal center.";

J. Biol. Inorg. Chem. 4:243-256(1999).

-!- FUNCTION: Ferritin is an intracellular molecule that stores iron

AC

DT

DT

DT

DE

GN

OS

OC

OC

OX

RN

RP

RX

RA

RA

RA

RL

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Q8R7U4;

28-FEB-2003 (Rel. 41, Created)

28-FEB-2003 (Rel. 41, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)

50S Ribosomal protein L10.

RPLJ OR TIE2304.

OS Thermoanaerobacter tengcongensis.

OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;

OC Thermoanaerobacteriaceae; Thermoanaerobacter.

NCBI\_TaxID=119072;

[1]

SEQUENCE FROM N.A.

STRAIN=MA4 / JCM 11007;

MEDLINE=2192816; PubMed=11997336;

Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.;

Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.;

Tan H., Chen R., Wang J., Yu J., Yang H.;

"A complete sequence of T. tengcongensis genome.";

Genome Res. 12:689-700(2002).

-!- SIMILARITY: Belongs to the L10P family of ribosomal proteins.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AE013173; AAM25445.1; -.

HAMAP; MF 00362; -; 1.

InterPro; IPR001790; Ribosomal L10.

InterPro; IPR002363; Ribosomal\_L10eub.

Pfam; PF00466; Ribosomal\_L10; 1.

PROSITE; PS01109; RIBOSOMAL\_L10; FALSE NEG.

PROSITE; PS01109; RIBOSOMAL\_L10; FALSE NEG.

Ribosomal protein; Complete proteome.

SEQUENCE 177 AA; 19591 MW; 50DDCF896EF6F4E8 CRC64;

-----

Query Match 95.0%; Score 19; DB 1; Length 177;

Best Local Similarity 100.0%; Pred. No. 3.1e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

-----

2 VAEF 5

12 VAEF 15

-----

RESULT 39

DSBB\_PASMU STANDARD; PRT; 178 AA.

AC Q9L6B3; P57804;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Disulfide bond formation protein B (Disulfide oxidoreductase).

GN DSBB OR PM0046.

OS Pasteurella multocida.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Pasteurella.

NCBI\_TaxID=747;

[1]

SEQUENCE FROM N.A.

Fuller T.E., Kennedy M.J., Lowery D.E.;

"Identification of Pasteurella multocida virulence genes in a

septicemic mouse model using signature-tagged mutagenesis.";

Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

[2]

SEQUENCE FROM N.A.

STRAIN=Pm70;

RX MEDLINE=21145866; PubMed=11248100;

RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;

"Complete genomic sequence of Pasteurella multocida Pm70.";

Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

-----

-!- FUNCTION: Required for disulfide bond formation in some periplasmic proteins. Acts by oxidizing the dSba protein (by similarity).

-!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (by similarity).

-!- SIMILARITY: Belongs to the dsbB family.

-----

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AF237925; AAF69411.1; --  
 HAMAP; MF 02286; -- 1.  
 InterPro: IPR003752; dsbB.  
 Pfam; PF02600; DsbB; 1.  
 Chaperone; Transmembrane; Complete proteome.  
 DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).  
 TRANSMEM 15 31 POTENTIAL.  
 DOMAIN 32 49 PERIPLASMIC (POTENTIAL).  
 TRANSMEM 50 65 POTENTIAL.  
 DOMAIN 66 72 CYTOPLASMIC (POTENTIAL).  
 TRANSMEM 73 90 POTENTIAL.  
 DOMAIN 91 145 PERIPLASMIC (POTENTIAL).  
 TRANSMEM 146 164 POTENTIAL.  
 TRANSMEM 165 177 CYTOPLASMIC (POTENTIAL).  
 DISULFID 41 44 REDOX-ACTIVE (BY SIMILARITY).  
 DISULFID 105 131 REDOX-ACTIVE (BY SIMILARITY).  
 SEQUENCE 178 AA; 20036 MW; D2C45BB73B31F0BC CRC64;

Query Match 95.0%; Score 19; DB 1; Length 178;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5  
 |||||  
 b 108 VAEF 111

ESULT 40  
 SM5 DROME STANDARD; PRT; 178 AA.  
 D ESMS DROME STANDARD; PRT; 178 AA.  
 C P13096; Q9VBI9;  
 01-JAN-1990 (Rel. 13, Created)  
 01-JAN-1990 (Rel. 13, Last sequence update)  
 15-MAR-2004 (Rel. 43, Last annotation update)  
 Enhancer of split m5 protein (E(spl)m5).  
 HLHMS OR CG6096  
 Drosophila melanogaster (Fruit fly).  
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=89231639; PubMed=2540957;  
 X Klamdt C., Knust E., Tietze K., Campos-Ortega J.A.;  
 T "Closely related transcripts encoded by the neurogenic gene complex  
 enhancer of split of Drosophila melanogaster.";  
 L EMBO J. 8:203-210(1989).  
 [2]  
 N SEQUENCE FROM N.A.  
 STRAIN=Berkley;  
 MEDLINE=20196006; PubMed=10731132;  
 X Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 A Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 A George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 A Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 Wan K.E., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Franckoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.S.M.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,  
 RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong P., Gorell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hosini D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jallmel B., Kalusi F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
 RA Kimmel M.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 [3]  
 RN WRPW MOTIF.  
 RP MEDLINE=95094252; PubMed=8001118;  
 RX Faroukh Z., Finley R.L. Jr., Kidd T., Mainwright S.M., Ingham P.W.,  
 RA Brent R., Ish-Horowicz D.;  
 RT "Groucho is required for Drosophila neurogenesis, segmentation, and  
 RT sex determination and interacts directly with hairy-related bHLH  
 RT proteins.";  
 RL Cell 79:805-815(1994).  
 CC -!- FUNCTION: Participates in the control of cell fate choice by  
 CC uncommitted neuroectodermal cells in the embryo. Transcriptional  
 CC repressor. Binds DNA on N-box motifs: 5'-CACNAG-3'.  
 CC -!- SUBUNIT: Transcription repression requires formation of a complex  
 CC with a co-repressor protein (Groucho). Forms homodimers.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- DEVELOPMENTAL STAGE: Expressed at the time when separation of  
 CC neural and epidermal precursors cells occurs. Mesectodermal  
 CC expression appears shortly before the onset of gastrulation.  
 CC -!- DOMAIN: The orange domain and the basic helix-loop-helix motif  
 CC mediate repression of specific transcriptional activators, such  
 CC as basic helix-loop-helix protein dimers.  
 CC -!- DOMAIN: The C-terminal WRPW motif is a transcriptional repression  
 CC domain necessary for the interaction with Groucho, a  
 CC transcriptional co-repressor recruited to specific target DNA by  
 CC hairy-related proteins.  
 CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
 CC -!- SIMILARITY: Contains 1 orange domain.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X16552; CA334552.1; --  
 EMBL; A5003754; XAF56552.1; --  
 PIR; S03629; S03629.

TRANSFAC; T01644; -.  
 FlyBase; FBgn0002631; HLHm5.  
 GO; GO:0005634; C:nucleus; IDA.  
 GO; GO:0003677; F:DNA binding; IDA.  
 InterPro; IPR001092; HLH\_basic.  
 InterPro; IPR003650; Orange.  
 Pfam; PF00010; HLH; 1.  
 SMART; SMC0353; HLH; 1.  
 SMART; SMC0511; ORANGE; 1.  
 PROSITE; PS00888; HLH; 1.  
 Differentiation; Neurogenesis; Nuclear protein; DNA-binding;  
 Transcription regulation; Repressor.  
 DNA\_BIND 19 33  
 DOMAIN 34 74 BASIC DOMAIN.  
 DOMAIN 88 129 HELIX-LOOP-HELIX MOTIF.  
 DOMAIN 175 178 ORANGE.  
 DOMAIN 175 178 WRPW MOTIF.  
 SEQUENCE 178 AA; 19223 MW; 19363D0F6043C84F CRC64;  
 Query Match 95.0%; Score 19; DB 1; Length 178;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 2 VAEF 5  
 ||||  
 45 VAEF 48

RESULT 41  
 18 ARCFU  
 18 ARCFU STANDARD; PRT; 178 AA.  
 029147;  
 16-OCT-2001 (Rel. 40, Created)  
 16-OCT-2001 (Rel. 40, Last sequence update)  
 16-OCT-2001 (Rel. 40, Last annotation update)  
 Hypothetical protein AF1118 precursor.  
 AF1118.  
 Archaeoglobus fulgidus.  
 Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
 Archaeoglobaceae; Archaeoglobus.  
 NCBI\_TaxID=2234;  
 [1] -  
 SEQUENCE FROM N.A.  
 STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 MEDLINE=98049343; PubMed=9389475;  
 Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 Richardson D.L., Kexlavage A.R., Graham D.E., Kyriakides N.C.,  
 Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
 Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 Venter J.C.;  
 "The complete genome sequence of the hyperthermophilic, sulphate-  
 reducing archaeon Archaeoglobus fulgidus";  
 Nature 390:364-370 (1997).  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; AE001027; AB090141.1; -.  
 TIGR; E69389; E69389.  
 TIGR; AF1118; -.  
 Hypothetical protein; Signal; Complete proteome.  
 SIGNAL 1 20 POTENTIAL  
 CHAIN 21 178 HYPOTHETICAL PROTEIN AF1118.

SQ SEQUENCE 178 AA; 19319 MW; 378A4F200240D924 CRC64;  
 Query Match 95.0%; Score 19; DB 1; Length 178;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 2 VAEF 5  
 ||||  
 155 VAEF 158

RESULT 42  
 APT\_HAEDU  
 ID APT\_HAEDU STANDARD; PRT; 179 AA.  
 AC Q7VKQ4;  
 DT 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).  
 GN APT OR HD1818.  
 OS Haemophilus ducreyi.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=730;  
 RN [1]  
 SEQUENCE FROM N.A.  
 STRAIN=3500HP / ATCC 700724;  
 RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,  
 Johnson L., Nguyen D., Wang J., Forst C., Hood L.;  
 "The complete genome sequence of Haemophilus ducreyi";  
 Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Catalyzes a salvage reaction resulting in the formation  
 of AMP, that is energetically less costly than de novo synthesis.  
 CC -!- CATALYTIC ACTIVITY: AMP + diphosphate = adenine + 5-phospho-alpha-  
 D-ribose 1-diphosphate.  
 CC -!- PATHWAY: Purine salvage.  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to the purine/pyrimidine  
 phosphoribosyltransferase family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; AE017156; AAP96568.1; -.  
 DR HAMAP; MF\_00004; -.  
 DR InterPro; IPR005764; Ade\_phospho\_trans.  
 DR InterPro; IPR002375; Pr/PY\_xp\_transf.  
 DR InterPro; IPR000836; PRtransferase.  
 DR Pfam; PF00156; PRibosyltran; 1.  
 DR TIGR; TIGR01090; apt; 1.  
 DR PROSITE; PS00103; PUR\_PYR\_PR\_TRANSFER; 1.  
 DR Transferase; Glycosyltransferase; Purine salvage; Complete proteome.  
 KW SEQUENCE 179 AA; 19315 MW; 13A1DAFDC13560DF CRC64;  
 Query Match 95.0%; Score 19; DB 1; Length 179;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 2 VAEF 5  
 ||||  
 45 VAEF 48

RESULT 43  
 RETB\_BOVIN  
 ID RETB\_BOVIN STANDARD; PRT; 183 AA.  
 AC P18902;

01-NOV-1990 (Rel. 16, Created).  
 01-NOV-1990 (Rel. 16, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 Plasma retinol-binding protein (PRBP) (RBP).  
 RBP4.  
 Bos taurus (Bovine).  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 NCBI\_TaxID=9913;  
 [1]  
 MEDLINE=91006139; PubMed=2209607;  
 Berni R., Stoppini M., Zapponi M.C., Meloni M.L., Monaco H.L.,  
 Zanotti G.;  
 "The bovine plasma retinol-binding protein. Amino acid sequence,  
 interaction with transthyretin, crystallization and preliminary X-ray  
 data."; Eur. J. Biochem. 192:507-513 (1990).  
 [2]  
 SEQUENCE OF 52-183 FROM N.A.  
 MEDLINE=93385352; PubMed=8373966;  
 Liu K.H., Dore J.J. Jr., Roberts M.P., Krishnan R., Hopkins F.M.,  
 Godkin J.D.;  
 "Expression and cellular localization of retinol-binding protein  
 messenger ribonucleic acid in bovine blastocysts and extraembryonic  
 membranes."; Biol. Reprod. 49:393-400 (1993).  
 [3]  
 X-RAY CRYSTALLOGRAPHY.  
 MEDLINE=92322903; PubMed=1623143;  
 Monaco H.L., Zanotti G.;  
 "three-dimensional structure and active site of three hydrophobic  
 molecule-binding proteins with significant amino acid sequence  
 similarity."; Biopolymers 32:457-465 (1992).  
 [4]  
 X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 MEDLINE=93266508; PubMed=8496140;  
 Zanotti G., Berni R., Monaco H.L.;  
 "Crystal structure of liganded and unliganded forms of bovine plasma  
 retinol-binding protein."; J. Biol. Chem. 268:10728-10738 (1993).  
 -1- FUNCTION: Delivers retinol from the liver stores to the peripheral  
 tissues. In plasma, the RBP-retinol complex interacts with  
 transthyretin, this prevents its loss by filtration through the  
 kidney glomeruli.  
 -1- SUBCELLULAR LOCATION: Secreted.  
 -1- SIMILARITY: Belongs to the lipocalin family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; S65585; AAB28336.1; --  
 PIR; I46355; I46955.  
 PIR; S13186; S13186.  
 PDB; IERB; 31-JAN-94.  
 PDB; IHBP; 31-JAN-94.  
 PDB; IHBO; 31-JAN-94.  
 PDB; IPEL; 01-NOV-94.  
 PDB; IPEM; 01-NOV-94.  
 PDB; IFEN; 01-NOV-94.  
 InterPro; IPR002345; Lipocalin.  
 InterPro; IPR000566; Lipocalin\_cytFABP.  
 Pfam; PF00061; Lipocalin; 1.  
 PRINTS; PR00179; LIPOCALIN.  
 PROSITE; PS00213; LIPOCALIN; 1.  
 Plasma; Vitamin A; Retinol-binding; Transport; Liver; Lipocalin;

3D-structure. 4 160 BY SIMILARITY.  
 FT DISULFID 70 174 BY SIMILARITY.  
 FT DISULFID 120 129 BY SIMILARITY.  
 FT HELIX 6 8  
 FT TURN 13 14  
 FT TURN 17 20  
 FT STRAND 22 30  
 FT TURN 33 34  
 FT STRAND 39 47  
 FT TURN 49 50  
 FT STRAND 53 62  
 FT TURN 64 65  
 FT STRAND 68 79  
 FT TURN 83 84  
 FT STRAND 85 92  
 FT TURN 95 96  
 FT STRAND 100 109  
 FT STRAND 114 123  
 FT TURN 125 126  
 FT STRAND 129 138  
 FT TURN 141 142  
 FT HELIX 146 158  
 FT TURN 159 160  
 FT TURN 162 163  
 FT STRAND 166 167  
 FT TURN 173 175  
 SQ SEQUENCE 183 AA; 21068 MW; D6BA064CB9E67C09 CRC64;  
 Query Match 95.0%; Score 19; DB 1; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 Db 42 VAEF 45  
 RESULT 44  
 DNAA WOLSP  
 ID DNAA WOLSP STANDARD; PRT; 186 AA.  
 AC P35907;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chromosomal replication initiator protein dnaA (Fragment).  
 DNAA.  
 GN Wolbachia sp.  
 OS Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
 OC Rickettsiaceae; Wolbachiae; Wolbachia.  
 OX NCBI\_TaxID=956;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95202115; PubMed=7894745;  
 RA Bourzis K., Nirgianaki A., Onyango P., Savakis C.;  
 RT "A prokaryotic dnaA sequence in *Drosophila melanogaster*: Wolbachia  
 infection and cytoplasmic incompatibility among laboratory strains."; Insect Mol. Biol. 3:131-142 (1994).  
 RL -1- FUNCTION: Plays an important role in the initiation and regulation  
 of chromosomal replication. Binds to the origin of replication; it  
 binds specifically double-stranded DNA at a 9 bp consensus (dnaA  
 box): 5'-TTATC(C/A)A(C/A)-3'. DnaA binds to ATP and to acidic  
 phospholipids (By similarity).  
 -1- SIMILARITY: Belongs to the dnaA family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

```

EMBL; Z28981; CAA82285.1; -.
PIR; S39317; S39317.
HAWAP; MF_00377; -.
InterPro; IPR001957; Bac_DnaA.
Pfam; PF00308; bac_dnaA.1.
PRINTS; PR00051; DNaA.
PROSITE; PS01008; DNaA; PARTIAL.
DNA replication; DNA-binding; ATP-binding.
NON_TER 1
NON_TER 186
SEQUENCE 186 AA; 21283 MW; C423C06CEFFB0459 CRC64;
Query Match 95.0%; Score 19; DB 1; Length 186;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2 VAEF 5
134 VAEF 137
|||||
PRT; 190 AA.
UCRI PARDE STANDARD; PRT; 190 AA.
P05417;
01-NOV-1988 (Rel. 09, Created)
01-NOV-1988 (Rel. 09, Last sequence update)
15-MAR-2004 (Rel. 43, Last annotation update)
Ubiquinol-cytochrome c reductase iron-sulfur subunit (EC 1.10.2.2)
(Rieske iron-sulfur protein) (RISP).
PEVA.
Paracoccus denitrificans.
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
Rhodobacteraceae; Paracoccus.
NCBI_TaxID=286;
[1]
SEQUENCE FROM N.A.
MEDLINE=88007612; PubMed=2820981;
Kurowski B., Ludwig B.
"The genes of the Paracoccus denitrificans bc1 complex. Nucleotide
sequence and homologies between bacterial and mitochondrial
subunits."
J. Biol. Chem. 262:13805-13811(1987).
-!- FUNCTION: Component of the ubiquinol-cytochrome c reductase
complex (complex III or cytochrome b-c1 complex), which is a
respiratory chain that generates an electrochemical potential
coupled to ATP synthesis.
-!- CATALYTIC ACTIVITY: 2H+ + 2 ferricytochrome c = Q + 2
ferricytochrome c.
-!- COFACTOR: Binds 1 2Fe-2S iron sulfur cluster per subunit (By
similarity).
-!- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
cytochrome c1 and the Rieske protein.
-!- SUBCELLULAR LOCATION: Membrane-bound.
-!- MISCELLANEOUS: The Rieske protein is a high potential 2Fe-2S
protein.
-!- SIMILARITY: Belongs to the Rieske family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M17522; AAA25571.1; -.
EMBL; X05799; CAA29243.1; -.
PIR; A29413; A29413.
HSPB; P13272; IRIE.
InterPro; IPR005805; Rieske.
InterPro; IPR005806; Rieske dom.
InterPro; IPR006317; Rieske_proteo.

```

```

DR InterPro; IPR006311; Tat.
DR Pfam; PF00355; Rieske; 1.
DR PRINTS; PRO0162; RIESKE.
DR TIGRFAMs; TIGR01416; Rieske_proteo; 1.
DR TIGRFAMs; TIGR01409; Tat_signal_seq; 1.
DR PROSITE; PS00199; RIESKE_1; 1.
DR PROSITE; PS00200; RIESKE_2; 1.
KW Electron transport; Inner membrane; Transmembrane; Metal-binding;
Iron-sulfur; Iron; 2Fe-2S; Oxidoreductase.
FT TRANSMEM 18 39 POTENTIAL.
FT METAL 132 132 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 134 134 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 152 152 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 155 155 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT DISULFID 137 154 BY SIMILARITY.
SQ SEQUENCE 190 AA; 20299 MW; F83F5D9A9C1FBEE4 CRC64;
Query Match 95.0%; Score 19; DB 1; Length 190;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2 VAEF 5
178 VAEF 181
|||||
PRT; 193 AA.
ID NTPA_METJA STANDARD; PRT; 193 AA.
AC Q57679;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Nucleoside-triphosphatase (EC 3.6.1.15) (Nucleoside triphosphate
phosphohydrolase) (NTPase).
DE MJ0226.
GN Methanococcus jannaschii.
OS Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=868087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kurlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geochagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii."
RL Science 273:1058-1073(1996).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=99332061; PubMed=10404228;
RA Hwang K.Y., Chung J.H., Kim S.-H., Han Y.S., Cho Y.;
RT "Structure-based identification of a novel NTPase from Methanococcus
jannaschii."
RL Nat. Struct. Biol. 6:691-696(1999).
CC -!- FUNCTION: CAN EFFICIENTLY HYDROLYZE NONSTANDARD NUCLEOTIDES SUCH
AS XTP TO XMP OR ITP TO IMP, BUT NOT THE STANDARD NUCLEOTIDES. XTP
IS THE BEST SUBSTRATE.
CC -!- CATALYTIC ACTIVITY: NTP + H(2)O = NDP + phosphate.
CC -!- COFACTOR: Manganese or manganese.
CC -!- SUBUNIT: Homodimer.
CC -!- SIMILARITY: Belongs to the HAM1 NTPase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
InterPro; IPR006317; Rieske_proteo.

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U67478; BAB98211.1; -  
 PIR; C64328; C64328  
 PDB; 1B78; 28-JAN-00.  
 PDB; 2MJP; 28-JAN-00.  
 TIGR; MJ0226; -  
 HAMAP; MF\_01405; atypical; 1.  
 InterPro; IPR002637; Hamip like.  
 Pfam; PF01725; Hamip like; 1.  
 TIGRFAMs; TIGR00042; TIGR00042; 1.  
 Hydrolase; Manganese; Magnesium; 3D-structure; Complete proteome.

STRAND 11 14  
 HELIX 18 27  
 TURN 28 30  
 TURN 32 33  
 TURN 36 39  
 STRAND 46 46  
 STRAND 50 65  
 STRAND 69 78  
 HELIX 79 81  
 TURN 82 83  
 STRAND 85 86  
 TURN 87 88  
 HELIX 89 95  
 TURN 96 96  
 HELIX 97 105  
 TURN 106 107  
 STRAND 112 123  
 TURN 124 125  
 STRAND 126 138  
 HELIX 150 153  
 STRAND 155 157  
 TURN 158 159  
 HELIX 164 166  
 HELIX 169 172  
 TURN 173 175  
 HELIX 177 190  
 TURN 191 191  
 SEQUENCE 193 AA; 22202 MW; 3570565E07D3DAC CRC64;

Query Match 95.0%; Score 19; DB 1; Length 193;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
 ||||  
 53 VAEF 56

RESULT 47  
 HAM2 BACHD STANDARD; PRT; 194 AA.  
 Q9K8D3;  
 28-FEB-2003 (Rel. 41, Created)  
 28-FEB-2003 (Rel. 41, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 HAM1 protein homolog 2.  
 BH3067  
 Bacillus halodurans.  
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 NCBI\_TaxID=86665;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=C-125 / JCM 9153;  
 MEDLINE=20512582; PubMed=11058132;  
 Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 Fuji F., Hirano C., Nakamura Y., Ogasawara N., Kuhara S.,  
 Horikoshi K.;

"Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis.";  
 Nucleic Acids Res. 28:4317-4331(2000).  
 -/- SIMILARITY: Belongs to the HAM1 NTPase family.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AF001517; BAB06786.1; -  
 PIR; C84033; C84033.  
 HSP; Q57679; 1B78.  
 HAMAP; MF\_01405; -; 1.  
 InterPro; IPR002637; Hamip like.  
 Pfam; PF01725; Hamip like; 1.  
 ProDom; PD004952; Hamip like; 1.  
 TIGRFAMs; TIGR00042; TIGR00042; 1.  
 Hydrolase; Complete proteome.  
 SEQUENCE 194 AA; 21868 MW; BBC5A0C4F19A04B3 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 194;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
 ||||  
 14 VAEF 17

RESULT 48  
 TNR0 ECOLI STANDARD; PRT; 194 AA.  
 AC P05823;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Transposon Tn2501 resolvase.  
 GN TNP.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 NCBI\_TaxID=562;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=87109049; PubMed=3027041;  
 Michiels T., Cornelis G., Ellis K., Grinstead J.;  
 "Tn2501, a component of the lactose transposon Tn951, is an example of a new category of class II transposable elements.";  
 J. Bacteriol. 169:624-631(1987).  
 -/- FUNCTION: RESOLVASE CATALYZES THE RESOLUTION (A SITE-SPECIFIC RECOMBINATION) OF THE COINTEGRATED REPLICON TO YIELD THE FINAL TRANSPOSITION PRODUCTS.  
 -/- MISCELLANEOUS: Tn2501 IS A CRYPTIC CLASS II TRANSPOSON FOUND AS PART OF THE LACTOSE TRANSPOSON Tn951.  
 -/- SIMILARITY: Belongs to the site-specific recombinase resolvase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M15197; AAA27426.1; -  
 PIR; B27758; RPECRS.  
 HSP; P03012; 2RSL.  
 InterPro; IPR006120; HTH\_7.

```

InterPro; IPR006118; Recombinase.
InterPro; IPR006119; resolvase_N.
Pfam; PF02796; HTH_7; 1.
Pfam; PF00239; resolvase; 1.
PROSITE; PS00397; RECOMBINASES_1; 1.
PROSITE; PS00398; RECOMBINASES_2; 1.
DNA recombination; DNA integration; DNA-binding; Transposable element.
ACT_SITE 11
TRANSIENT COVALENT LINKAGE TO DNA DURING
STRAND CLEAVAGE AND REJOINING
(BY SIMILARITY)
DNA BIND 170 189
H-T-H MOTIF (PROBABLE).
SEQUENCE 194 AA; 21420 MW; DEDFC0BA6406EAD1 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 194;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5
|||||
120 VAEF 123

>PINO ECOLI STANDARD; PRT; 196 AA.
P77170;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Putative DNA-invertase from lambdaoid prophage Qin.
PINO OR B1545.
Escherichia coli.
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
NCBI_TaxID=562;
[1]
SEQUENCE FROM N.A.
STRAIN=K12 / MG1655;
MEDLINE=97251357; PubMed=9278503;
Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
"The complete genome sequence of Escherichia coli K-12."
Science 277:1453-1474(1997).
[2]
SEQUENCE FROM N.A.
STRAIN=K12;
MEDLINE=97251357; PubMed=9097039;
Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
"A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map."
DNA Res. 3:363-377(1996).
-!- SIMILARITY: Belongs to the site-specific recombinase resolvase
family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; AB000252; AAC74618.1; -.
EMBL; D90798; BAA15249.1; -.
PIR; D64909; D64909.
HSSP; P03012; 2RSL.

```

```

EcGene; EG13824; pinQ.
DR InterPro; IPR006120; HTH_7.
DR InterPro; IPR006118; Recombinase.
DR InterPro; IPR006119; resolvase_N.
DR Pfam; PF02796; HTH_7; 1.
DR Pfam; PF00239; resolvase; 1.
DR PROSITE; PS00397; RECOMBINASES_1; FALSE_NEG.
DR PROSITE; PS00398; RECOMBINASES_2; 1.
KW Hypothetical protein; DNA recombination; DNA integration; DNA-binding;
KW DNA invertase; Complete proteome.
FT ACT_SITE 11
11 TRANSIENT COVALENT LINKAGE TO DNA DURING
STRAND CLEAVAGE AND REJOINING
(BY SIMILARITY)
FT SEQUENCE 196 AA; 21852 MW; 7CF47F6D98DBD497 CRC64;
SQ

Query Match 95.0%; Score 19; DB 1; Length 196;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5
|||||
120 VAEF 123

>PINO ECOLI STANDARD; PRT; 196 AA.
P77574;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Putative DNA-invertase from lambdaoid prophage Rac.
PINO OR B1374 OR C3146.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992;
[1]
SEQUENCE FROM N.A.
STRAIN=K12 / MG1655;
MEDLINE=97426617; PubMed=9278503;
Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
"The complete genome sequence of Escherichia coli K-12."
Science 277:1453-1474(1997).
[2]
SEQUENCE FROM N.A.
STRAIN=K12;
MEDLINE=97251357; PubMed=9097039;
Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
"A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map."
DNA Res. 3:363-377(1996).
[3]
SEQUENCE FROM N.A.
STRAIN=O6:HI / CFT073 / ATCC 700928;
MEDLINE=22388234; PubMed=12471157;
Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
Moley H.L.T., Donnenberg M.S., Blattner F.R.;
"Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli."
Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
-!- SIMILARITY: Belongs to the site-specific recombinase resolvase

```

```

C family.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
C R EMBL; AE000234; AAC74456.1; -.
C R EMBL; D90775; BAA14979.1; -.
C R EMBL; AE016764; AAN81596.1; -.
C R PIR; A64888; A64888.
C R HSP; P03012; 2RSL.
C R EcGene; EGI3372; pInR.
C R InterPro; IPR006120; HTH 7.
C R InterPro; IPR006118; Recombinase.
C R InterPro; IPR006119; resolvase_N.
C R Pfam; PFO2796; HTH 7; 1.
C R Pfam; PFO0239; resolvase; 1.
C R PROSITE; PS00397; RECOMBINASES_1; FALSE_NEG.
C R PROSITE; PS00398; RECOMBINASES_2; 1.
C R Hypothetical protein; DNA recombination; DNA integration; DNA-binding;
C W DNA invertase; Complete proteome.
C T ACT_SITE 11 11
C T TRANSIENT COVALENT LINKAGE TO DNA DURING
C T STRAND CLEAVAGE AND REJOINING
C T (BY SIMILARITY).
C T SEQUENCE 196 AA; 21908 MW; 643A62BC940B29A9 CRC64;
C
C Query Match 95.0%; Score 19; DB 1; Length 196;
C Best Local Similarity 100.0%; Pred. No. 3.4e+02;
C Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
C
C Y 2 VAEF 5
C b 120 VAEF 123
C
C ESULT 51
C HPA_ECOLI
C D_UHPA_ECOLI STANDARD; PRT; 196 AA.
C
C I P10940;
C I 01-JUL-1989 (Rel. 11, Created)
C I 01-JUL-1989 (Rel. 11, Last sequence update)
C I 10-OCT-2003 (Rel. 42, Last annotation update)
C E Transcriptional regulatory protein uhpA.
C UHPA OR B3669 OR C4593 OR Z5159 OR ECS4606 OR SF3792 OR S3976.
C Escherichia coli,
C S Escherichia coli O6,
C S Escherichia coli O157:H7, and
C S Shigella flexneri.
C S Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C S Enterobacteriaceae; Escherichia.
C X NCBI_TaxID=562, 217992, 83334, 623;
C N [1]
C N SEQUENCE FROM N.A.
C P SPECIES=E.coli;
C X MEDLINE=87279903; PubMed=3301805;
C X Friedrich M.J., Kadner R.J.;
C T "Nucleotide sequence of the uhp region of Escherichia coli.";
C J. Bacteriol. 169:3556-3563(1987).
C [2]
C N SEQUENCE FROM N.A.
C P SPECIES=E.coli;
C X MEDLINE=92234930; PubMed=1569007;
C X Island M.D., Wei B.-Y., Kadner R.J.;
C T "Structure and function of the uhp genes for the sugar phosphate
C T transport system in Escherichia coli and Salmonella typhimurium.";
C J. Bacteriol. 174:2754-2762(1992).
C [3]
C N SEQUENCE FROM N.A.
C P SPECIES=E.coli; STRAIN=K12 / MGL655;

```

```

RX MEDLINE=93315143; PubMed=7686882;
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication.";
RL Genomics 16:551-561(1993).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [7]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao Y., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [8]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
CC -!- FUNCTION: APPEARS TO BE THE POSITIVE ACTIVATOR OF UHP
CC TRANSCRIPTION.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- PTM: PHOSPHORYLATED BY UHPB (PROBABLE).
CC -!- SIMILARITY: Contains 1 response regulatory domain.
CC -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```



Davidson N.O., Honjo T.;  
 "Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells.";  
 J. Biol. Chem. 274:18470-18476(1999).  
 -!- FUNCTION: RNA-editing deaminase involved in somatic hypermutation, gene conversion, and class-switch recombination. Required for several crucial steps of B-cell terminal differentiation necessary for efficient antibody responses.  
 -!- CATALYTIC ACTIVITY: Cytidine + H(2)O = uridine + NH(3).  
 -!- COFACTOR: Zinc (By similarity).  
 -!- SIMILARITY: Belongs to the cytidine and deoxycytidylate deaminases family.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; AF132979; AAD41793.1; .  
 GO; MGI:1342279; Aicda.  
 GO; GO:0004126; F:cytidine deaminase activity; IDA.  
 InterPro; IPR007904; APOBEC\_C.  
 InterPro; IPR002125; dCMP/cyt\_deam.  
 Pfam; PF05240; APOBEC\_C; 1.  
 PROSITE; PS00903; CYT\_DCMF\_DEAMINASES; 1.  
 mRNA processing; Hydroxylase; Zinc.  
 METAL 56 56 ZINC (BY SIMILARITY).  
 METAL 87 87 ZINC (BY SIMILARITY).  
 METAL 90 90 ZINC (BY SIMILARITY).  
 SEQUENCE 198 AA; 24030 MW; 18A3BA10CA54BEB2 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 198;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
 ||||  
 94 VAEF 97

-----  
 RESULT 54  
 US5\_SULSO  
 D\_HIS5\_SULSO STANDARD; PRT; 199 AA.  
 C 033777;  
 15-DEC-1998 (Rel. 37, Created)  
 15-DEC-1998 (Rel. 37, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 Imidazole glycerol phosphate synthase subunit hisH (EC 2.4.2.-) (IGP synthase glutamine amidotransferase subunit) (IGP synthase subunit hisH) (IMP synthase subunit hisH) (IGPS subunit hisH).  
 HIS (HIS OR SSO0600 OR C08\_050).  
 Sulfolobus solfataricus.  
 Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae; Sulfolobus.  
 NCBI\_TaxID=2287;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=ATCC 35092 / DSM 1617 / P2;  
 MEDLINE=97352708; PubMed=9209067;  
 Charlebois R.L., Sensen C.W., Doolittle W.F., Brown J.R.;  
 "Evolutionary analysis of the hisCABDpDEHI gene cluster from the archaeson Sulfolobus solfataricus P2.";  
 J. Bacteriol. 179:4429-4432(1997).  
 [2]  
 SEQUENCE FROM N.A.  
 STRAIN=ATCC 35092 / DSM 1617 / P2;  
 MEDLINE=20165948; PubMed=10701121;  
 Charlebois R.L., Singh R.K., Chan-Weiher C.C.-Y., Allard G., Chow C., Confalonieri F., Curtis B., Duguet M., Erauso G., Faguy D.,

GAasterland T., Garrett R.A., Gordon P., Jeffries A.C., Kozera C., Kushwaha N., Lafluer E., Medina N., Peng X., Penny S.L., She Q., St Jean A., van der Oost J., Young F., Zivanovic Y., Doolittle W.F., Ragan M.A., Sensen C.W.;  
 "Gene content and organization of a 281-kbp contig from the genome of the extremely thermophilic archaeson, Sulfolobus solfataricus P2.";  
 Genome 43:116-136(2000).  
 [3]  
 SEQUENCE FROM N.A.  
 STRAIN=ATCC 35092 / DSM 1617 / P2;  
 MEDLINE=21332296; PubMed=11427726;  
 She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G., Aveyez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A., De Moers A., Erauso G., Fletcher C., Gordon P.M.K., Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X., Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N., Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T., Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;  
 "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 -!- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to IGP, AICAR and glutamate. The hisH subunit provides the glutamine amidotransferase activity that produces the ammonia necessary to hisP for the synthesis of IGP and AICAR (By similarity).  
 -!- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribose-1-ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.  
 -!- PATHWAY: Histidine biosynthesis; fifth step.  
 -!- SUBUNIT: Heterodimer of hisH and hisP (By similarity).  
 -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; U82227; AAB63025.1; .  
 EMBL; Y18930; CAB57700.1; .  
 EMBL; AEO06689; AAK40911.1; .  
 PIR; H90206; H90206.  
 HAMAP; MF\_00278; -; 1.  
 InterPro; IPR000991; GATase\_1.  
 Pfam; PF00117; GATase\_1.  
 PROSITE; PS00442; GATASE\_TYPE1; FALSE NEG.  
 Histidine biosynthesis; Transferase; Glutamine amidotransferase; Complete proteome.  
 ACT\_SITE 76 76 BY SIMILARITY.  
 ACT\_SITE 178 178 BY SIMILARITY.  
 ACT\_SITE 180 180 BY SIMILARITY.  
 SEQUENCE 199 AA; 22533 MW; 17DB1A86724CE095 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
 ||||  
 51 VAEF 54

-----  
 RESULT 55  
 RETB\_HORSE  
 ID\_RETB\_HORSE STANDARD; PRT; 201 AA.  
 AC Q28359;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Plasma retinol-binding protein precursor (PRBP) (RBP).

RBPA.  
Equus caballus (Horse).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
NCBI\_TaxID=9796;  
[1]  
SEQUENCE FROM N.A.  
TISSUE=Endometrium;  
MEDLINE=95226610; PubMed=7536053;  
McDowell K.J., Adams M.H., Franklin K.M., Baker C.B.;  
"Changes in equine endometrial retinol-binding protein RNA during the  
estrous cycle and early pregnancy and with exogenous steroids.";  
Biol. Reprod. 52:438-443(1995).  
-!- FUNCTION: Delivers retinol from the liver stores to the peripheral  
tissues. In plasma, the RBP-retinol complex interacts with  
transferrin, this prevents its loss by filtration through the  
kidney glomeruli.  
-!- SUBCELLULAR LOCATION: Secreted.  
-!- SIMILARITY: Belongs to the lipocalin family.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U21208; AAC4861.1; --  
PIR; I46257; I46257.  
HSP; P27485; IAQB.  
InterPro; IPR002345; lipocalin.  
InterPro; IPR005666; lipocalin\_cytFABP.  
Pfam; PF00061; lipocalin; 1.  
PRINTS; PR00179; LIPOCALIN.  
PROSITE; PS00213; LIPOCALIN; 1.  
Plasma; vitamin A; Retinol-binding; Transport; Liver; Signal;  
Lipocalin.  
SIGNAL 1 18 POTENTIAL.  
CHAIN 19 201 PLASMA RETINOL-BINDING PROTEIN.  
DISULFID 22 178 BY SIMILARITY.  
DISULFID 88 192 BY SIMILARITY.  
DISULFID 138 147 BY SIMILARITY.  
SEQUENCE 201 AA; 23022 MW; 12CF90834E4262DC CRC64;  
Query Match 95.0%; Score 19; DB 1; Length 201;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
2 VAEF 5  
60 VAEF 63  
SULT 56  
TB HUMAN  
RETE HUMAN STANDARD; PRT; 201 AA.  
P02753; O43478; O43479; Q8WAA3; Q9P178;  
21-JUL-1986 (Rel. 01, Created)  
28-FEB-2003 (Rel. 41, Last sequence update)  
10-OCT-2003 (Rel. 42, Last annotation update)  
Plasma retinol-binding protein precursor (PRBP) (RBP) (PRO2222).  
RBPA.  
Homo sapiens (Human).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=84069802; PubMed=6316270;  
Colantuoni V., Romano V., Bensi G., Santoro C., Costanzo F.,  
Raugei G., Cortese R.;  
"Cloning and sequencing of a full length cDNA coding for human  
retinol-binding protein.";  
Nucleic Acids Res. 11:7769-7776(1983).  
[2]  
SEQUENCE FROM N.A.  
TISSUE=Liver;  
MEDLINE=22388257; PubMed=12477932;  
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
Besak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[3]  
SEQUENCE OF 1-189 FROM N.A.  
MEDLINE=86055755; PubMed=29987779;  
D'Onofrio C., Colantuoni V., Cortese R.;  
"Structure and cell-specific expression of a cloned human retinol  
binding protein gene: the 5'-flanking region contains hepatoma  
specific transcriptional signals.";  
EMBO J. 4:1981-1989(1985).  
[4]  
SEQUENCE OF 19-201, AND DISULFIDE BONDS.  
MEDLINE=86019004; PubMed=24440324;  
Raak L., Anundi H., Fohlman J., Peterson P.A.;  
"The complete amino acid sequence of human serum retinol-binding  
protein.";  
Ups. J. Med. Sci. 92:115-146(1987).  
[5]  
SEQUENCE OF 19-201.  
MEDLINE=81254137; PubMed=6942701;  
Raak L., Anundi H., Boehme J., Eriksson U., Ronne H., Sege K.,  
Peterson P.A.;  
"Structural and functional studies of vitamin A-binding proteins.";  
Ann. N.Y. Acad. Sci. 359:79-90(1981).  
[6]  
SEQUENCE OF 19-183.  
MEDLINE=80004132; PubMed=573217;  
Raak L., Anundi H., Peterson P.A.;  
"The primary structure of the human retinol-binding protein.";  
FEBS Lett. 104:55-58(1979).  
[7]  
SEQUENCE OF 18-201 FROM N.A.  
TISSUE=Fetal liver;  
Zhang C., Yu Y., Zhang S., Wei H., Zhang Y., Zhou G., Bi J., Liu M.,  
He F.;  
"Functional prediction of the coding sequences of 79 new genes deduced  
by analysis of cDNA clones from human fetal liver.";  
Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
[8]  
PARTIAL SEQUENCE.  
MEDLINE=95395382; PubMed=7666002;  
Jaconi S., Rose K., Hughes G.J., Saurat J.-H., Siegenthaler G.;  
"Characterization of two post-translationally processed forms of  
human serum retinol-binding protein: altered ratios in chronic renal  
failure.";  
J. Lipid Res. 36:1247-1253(1995).  
[9]  
X-RAY CRYSTALLOGRAPHY (2.0 ÅNGSTROMS).  
MEDLINE=91017498; PubMed=2217163;  
RX

Cowan S.W., Newcomer M.E., Jones T.A.;  
 "Crystallographic refinement of human serum retinol binding protein  
 at 2-A resolution";  
 Proteins 8:44-61(1990).  
 [10]  
 N P X-RAY CRYSTALLOGRAPHY.  
 MEDLINE=92322903; PubMed=1623143;  
 Monaco H.L., Zanotti G.;  
 "Three-dimensional structure and active site of three hydrophobic  
 molecule-binding proteins with significant amino acid sequence  
 similarity";  
 Biopolymers 32:457-465(1992).  
 [11]  
 N P X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS) OF COMPLEX WITH TTR.  
 MEDLINE=99162254; PubMed=10052934;  
 Naylor H.M., Newcomer M.E.;  
 "The structure of human retinol-binding protein (RBP) with its  
 carrier protein transthyretin reveals an interaction with the carboxy  
 terminus of RBP";  
 Biochemistry 38:2647-2653(1999).  
 [12]  
 N P VARIANTS RBP DEFICIENCY ASN-59 AND ASP-93.  
 MEDLINE=99103495; PubMed=9888420;  
 Seeliger M.W., Biesalski H.K., Wissingner B., Gollnick H., Gielen S.,  
 Frank J., Beck S., Zrenner E.;  
 "Phenotype in retinol deficiency due to a hereditary defect in retinol  
 binding protein synthesis";  
 Invest. Ophthalmol. Vis. Sci. 40:3-11(1999).  
 [13]  
 N P CHARACTERIZATION OF VARIANTS ASN-59 AND ASP-93.  
 MEDLINE=99247593; PubMed=10232633;  
 Biesalski H.K., Frank J., Beck S.C., Heinrich F., Illek B., Reifen R.,  
 Gollnick H., Seeliger M.W., Wissingner B., Zrenner E.;  
 "Biochemical but not clinical vitamin A deficiency results from  
 mutations in the gene for retinol binding protein";  
 Am. J. Clin. Nutr. 69:931-936(1999).  
 N P FUNCTION: Delivers retinol from the liver stores to the peripheral  
 tissues. In plasma, the RBP-retinol complex interacts with  
 transthyretin, this prevents its loss by filtration through the  
 kidney glomeruli.  
 N P SUBCELLULAR LOCATION: Secreted.  
 N P MASS SPECTROMETRY: MW=21063.46; MW ERR=1.88; METHOD=Electrospray;  
 RANGE=17-199.  
 N P DISEASE: Defects in RBP4 are a cause of retinol-binding protein  
 deficiency [MIM:180250]. This condition causes night vision  
 problems. It produces a typical "fundus xerophthalmicus,"  
 featuring a progressed atrophy of the retinal pigment epithelium.  
 N P DISEASE: A deficiency of vitamin A blocks secretion of the binding  
 protein posttranslationally and results in defective delivery and  
 supply of vitamin to the epidermal cells (a condition associated  
 with a dermatosis).  
 N P SIMILARITY: Belongs to the lipocalin family.  
 N P DATABASE: NAME-Mutations of the RBP4 gene;  
 NOTE=Retina International's Scientific Newsletter;  
 WWW="http://www.retina-international.com/sci-news/rbp4mut.htm".  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; X00129; CAA24959.1; -;  
 EMBL; BC020633; AAH20633.1; -;  
 EMBL; X02775; CAA24553.1; -;  
 EMBL; X02824; CAA6489.1; -;  
 EMBL; AF119868; AAF69622.1; ALT\_INIT.  
 EMBL; AF025334; AAC02945.1; -;  
 EMBL; AF025335; AAC02946.1; -;  
 PIR; A93494; VAHU.  
 PDB; 1BRP; 15-JUL-91.

DR PDB; 1BRP; 31-JAN-94.  
 DR PDB; 1BRQ; 31-JAN-94.  
 DR PDB; 1QAB; 09-APR-99.  
 DR SWISS-2DPAGE; P02753; HUMAN.  
 DR Sjiena-2DPAGE; P02753; -;  
 DR Genew; HGNC:9222; RBP4.  
 DR MIM; 180250; -; Cextracellular space; TAS.  
 DR GO; GO:0008615; F:retinoid binding; TAS.  
 DR GO; GO:0005501; F:retinoid binding; TAS.  
 DR InterPro; IPR002345; Lipocalin.  
 DR InterPro; IPR000566; Lipocalin\_cytFAPP.  
 DR Pfam; PF00061; Lipocalin; 1.  
 DR PRINTS; PR00179; LIPOCALIN.  
 DR PROSITE; PS00213; LIPOCALIN; 1.  
 DR Plasma; vitamin A; Retinol-binding; Transport; Liver; Signal;  
 KW Lipocalin; Disease mutation; Vision; 3D-structure.  
 KW SIGNAL 1 18 PLASMA RETINOL-BINDING PROTEIN.  
 FT CHAIN 19 201  
 FT DISULFID 22 178  
 FT DISULFID 88 192  
 FT DISULFID 138 147  
 FT VARIANT 59 59  
 FT VARIANT 93 93  
 FT VARIANT 93 93  
 FT CONFLICT 8 8  
 FT CONFLICT 13 17  
 FT STRAND 23 23  
 FT HELIX 24 26  
 FT TURN 31 32  
 FT HELIX 35 38  
 FT STRAND 40 48  
 FT STRAND 60 66  
 FT TURN 67 69  
 FT STRAND 70 83  
 FT TURN 82 83  
 FT STRAND 86 96  
 Query Match 95.0%; Score 19; DB 1; Length 201;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 VAEF 5  
 Db 60 VAEF 63  
 RESULT 57  
 RETB\_PIG ID RETB\_PIG STANDARD; PRT; 201 AA.  
 AC P27485;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Plasma retinol-binding protein precursor (PRBP) (RBP).  
 GN RBP4.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 CX NCBI\_TaxID=9823;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=92131004; PubMed=1723146;  
 RA Trout W.B., McDonnell J.J., Kramer K.K., Baumbach G.A., Roberts R.M.;  
 "The retinol-binding protein of the expanding pig blastocyst;  
 molecular cloning and expression in trophectoderm and embryonic  
 disc";  
 Mol. Endocrinol. 5:1533-1540(1991).  
 [2]  
 X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS), AND REVISIONS TO 134 AND 185.  
 MEDLINE=98437649; PubMed=9757135;  
 RA Zanotti G., Panzalcorto M., Marcato A., Malpeli G., Folli C., Berni R.;  
 "Structure of pig plasma retinol-binding protein at 1.65-A

resolution.;  
Acta Crystallogr. D 54:1049-1052(1998).  
-!- FUNCTION: Delivers retinol from the liver stores to the peripheral tissues. In plasma, the RBP-retinol complex interacts with kidney glomeruli.  
-!- SUBCELLULAR LOCATION: Secreted.  
-!- SIMILARITY: Belongs to the lipocalin family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M68860; AAA31113.1; -.  
PIR; A39486; A39486.  
PDB; 1AOB; 28-JAN-98.  
InterPro; IPR002345; Lipocalin.  
InterPro; IPR000566; Lipocalin\_cyFABP.  
Pfam; PF00061; Lipocalin; 1.  
PRINTS; PR00179; LIPOCALIN.  
PROSITE; PS00213; LIPOCALIN; 1.  
Plasma; Vitamin A; Retinol-binding; Transport; Liver; Signal; Lipocalin; 3D-structure.  
SIGNAL 1 18 POTENTIAL.  
CHAIN 19 201 PLASMA RETINOL-BINDING PROTEIN.  
DISULFID 22 178  
DISULFID 88 192  
DISULFID 138 147  
CONFLICT 134 134 V -> A (IN REF. 1).  
CONFLICT 185 185 I -> L (IN REF. 1).  
STRAND 23 23  
HELIX 24 26  
TURN 31 32  
STRAND 35 38  
HELIX 40 48  
STRAND 55 65  
TURN 67 68  
STRAND 71 80  
TURN 82 83  
STRAND 86 97  
TURN 101 102  
STRAND 103 110  
TURN 113 114  
STRAND 118 127  
STRAND 132 141  
TURN 143 144  
STRAND 146 156  
TURN 159 160  
HELIX 164 176  
TURN 177 178  
TURN 180 181  
STRAND 184 185  
TURN 191 192  
SEQUENCE 201 AA; 23067 MW; A20E39D3C9471DC8 CRC64;  
Query Match 95.0%; Score 19; DB 1; Length 201;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
2 VAEF 5  
60 VAEF 63  
SULT 59  
TB RABIT  
P06912;  
01-JAN-1988 (Rel. 06, Created)  
PRT; 201 AA.  
STANDARD;  
RETS RABIT  
STANDARD;  
PRT; 201 AA.  
STANDARD;  
PRT; 204 AA.  
STANDARD;  
PRT; 204 AA.

DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DE 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Plasma retinol-binding protein precursor (PRBP) (RBP).  
GN RBP4.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93011736; PubMed=1339354;  
RA Sundelin J., Laurent B.C., Anundi H., Traegaardh L., Larhammar D., Bjoerck L., Eriksson U., Akerstrom B., Jones A., Newcomer M., Peterson P.A., Rask L.;  
RA "Amino acid sequence homologies between rabbit, rat, and human serum retinol-binding proteins";  
RT J. Biol. Chem. 260:6472-6480(1985).  
RL -!- FUNCTION: Delivers retinol from the liver stores to the peripheral tissues. In plasma, the RBP-retinol complex interacts with transthyretin, this prevents its loss by filtration through the kidney glomeruli.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: Belongs to the lipocalin family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; S45958; AAB23582.1; -.  
PIR; A49178; VABP.  
HSP; P18902; IFEN.  
InterPro; IPR002345; Lipocalin.  
InterPro; IPR000566; Lipocalin\_cyFABP.  
Pfam; PF00061; Lipocalin; 1.  
PRINTS; PR00179; LIPOCALIN.  
PROSITE; PS00213; LIPOCALIN; 1.  
Plasma; Vitamin A; Retinol-binding; Transport; Liver; Lipocalin; Signal.  
SIGNAL 1 18 PLASMA RETINOL-BINDING PROTEIN.  
CHAIN 19 201  
DISULFID 22 178 BY SIMILARITY.  
DISULFID 88 192 BY SIMILARITY.  
DISULFID 138 147 BY SIMILARITY.  
SQ SEQUENCE 201 AA; 23102 MW; 4153FCF050184136 CRC64;  
Query Match 95.0%; Score 19; DB 1; Length 201;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 VAEF 5  
DB 60 VAEF 63  
RESULT 59  
YIGZ\_ECOLI  
ID YIGZ\_ECOLI STANDARD; PRT; 204 AA.  
AC P27852;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein yigZ.  
GN YIGZ OR B3848.

S Escherichia coli.  
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
C Enterobacteriaceae; Escherichia.  
X NCBI\_TaxID=562;  
N [1]  
P SEQUENCE FROM N.A.  
C STRAIN=K12 / MGL655;  
X MEDLINE=92359234; PubMed=1379743;  
A Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;  
T "Analysis of the Escherichia coli genome: DNA sequence of the region  
T from 84.5 to 85.5 minutes";  
L Science 257:771-778(1992).  
N [2]  
P SEQUENCE FROM N.A.  
C STRAIN=K12;  
X MEDLINE=91057145; PubMed=22433799;  
A Nakahigashi K., Inokuchi H.;  
T "Nucleotide sequence between the fadB gene and the rna operon from  
T Escherichia coli";  
L Nucleic Acids Res. 18:6439-6439(1990).  
N [3]  
P SMILARITY.  
X MEDLINE=94147996; PubMed=8313894;  
A Koonin E.V., Bork P., Sander C.;  
T "Yeast chromosome III: new gene functions.";  
L EMBO J. 13:493-503(1994).  
C -1- SIMILARITY: BELONGS TO THE UPF0029 FAMILY. STRONG, TO H.INFLUENZAE  
C HI0722.  
C This SWISS-PROT entry is copyright. It is produced through a collaboration  
C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
C the European Bioinformatics Institute. There are no restrictions on its  
C use by non-profit institutions as long as its content is in no way  
C modified and this statement is not removed. Usage by and for commercial  
C entities requires a license agreement (See http://www.isb-sib.ch/announce/  
C or send an email to license@isb-sib.ch).  
R EMBL; M87049; AAA67645.1; ALT INIT.  
R EMBL; AE000460; AAC76851.1; ALT INIT.  
R EMBL; X54687; CAA38501.1; ALT\_INIT.  
R Ecogene; EG11484; Y192.  
R InterPro; IPR009022; EFG\_III\_V.  
R InterPro; IPR001498; UPF0029.  
R Pfam; PF01205; UPF0029; 1.  
R TIGRPFAMs; TIGR00257; TIGR00257; 1.  
R PROSITE; PS00910; UPF0029; 1.  
W Hypothetical protein; Complete proteome.  
T CONFLICT 125 125 L -> V (IN REF. 2).  
Q SEQUENCE 204 AA; 21757 MW; 97E48163FB899DBE CRC64;

Query Match 95.0%; Score 19; DB 1; Length 204;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5  
b 182 VAEF 185

RESULT 60  
ID21\_HUMAN  
D MD21\_HUMAN STANDARD; PRT; 205 AA.  
C Q13257;  
T 16-OCT-2001 (Rel. 40, Created)  
T 16-OCT-2001 (Rel. 40, Last sequence update)  
T 10-OCT-2003 (Rel. 42, Last annotation update)  
E Mitotic spindle assembly checkpoint protein MAD2A (MAD2-like 1)  
E (HsMAD2).  
N MAD2L1 OR MAD2.  
N Homo sapiens (Human).  
S Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
C Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
X NCBI\_TaxID=9606;

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96421709; PubMed=8824189;  
RA Li Y., Benezra R.;  
RT "Identification of a human mitotic checkpoint gene: hsmad2.";  
RL Science 274:246-248(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX Jin D.-Y., Jeang K.-T.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBSJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX Klebert S., Barnikol-Watanabe S., Kratzin H.D., Hilschmann N.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBSJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RX Nobori T.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBSJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Bone marrow, and Muscle;  
RL MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Vax S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Souffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.N.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.B.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [6]  
RP INTERACTION WITH CDC20.  
RX MEDLINE=98301442; PubMed=9637688;  
RA Pang G., Yu H., Kirschner M.W.;  
RT "The checkpoint protein MAD2 and the mitotic regulator CDC20 form a  
RT ternary complex with the anaphase-promoting complex to control  
RT anaphase initiation.";  
RL Genes Dev. 12:1871-1883(1998).  
RN [7]  
RP INTERACTION WITH ADAM17.  
RX MEDLINE=99458684; PubMed=10527948;  
RA Nelson K.K., Schlondorff J., Blobel C.P.;  
RT "Evidence for an interaction of the metalloprotease-disintegrin tumour  
RT necrosis factor alpha convertase (TACE) with mitotic arrest deficient  
RT 2 (MAD2), and of the metalloprotease-disintegrin MDC9 with a novel  
RT MAD2-related protein, MAD2-beta.";  
RL Biochem. J. 343:673-680(1999).  
RN [8]  
RP STRUCTURE BY NMR OF 11-195.  
RX MEDLINE=20165182; PubMed=10700282;  
RA Luo X., Fang G., Coldiron M., Lin Y., Yu H., Kirschner M.W.,  
RA Wagner G.;  
RT "Structure of the Mad2 spindle assembly checkpoint protein and its  
RT interaction with Cdc20.";  
RL Nat. Struct. Biol. 7:224-229(2000).  
CC -1- FUNCTION: REQUIRED FOR THE EXECUTION OF THE MITOTIC CHECKPOINT  
CC WHICH MONITORS THE PROCESS OF KINETOCORE-SPINDLE ATTACHMENT AND  
CC DELAYS THE ONSET OF ANAPHASE WHEN THIS PROCESS IS NOT COMPLETE. IT  
CC INHIBITS THE ACTIVITY OF THE ANAPHASE PROMOTING COMPLEX BY  
CC SEQUESTERING CDC20 UNTIL ALL CHROMOSOMES ARE ALIGNED AT THE  
CC METAPHASE PLATE.

-----  
-1- SUBUNIT: INTERACTS WITH CDC20 AND WITH ADAM17 (TACE).  
-1- SUBCELLULAR LOCATION: Nuclear.  
-1- SIMILARITY: BELONGS TO THE MAD2 FAMILY.  
-1- SIMILARITY: Contains 1 HORVA domain.  
-1- DATABASE: NAME-Atlas Genet. Cytogenet. Oncol. Haematol.;  
WWW="http://www.infobiogen.fr/services/chronocancer/Genes/MAD2L1ID304.html".  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U65410; AAC50781.1; -;  
EMBL; U31278; AAC52060.1; -;  
EMBL; AJ000186; CAA03943.1; -;  
EMBL; AB056160; BAB63410.1; -;  
EMBL; BC000356; AAH00356.1; -;  
EMBL; BC005945; AAH05945.1; -;  
PIR; G01942; G01942;  
PDB; 1DUU; 08-MAR-00.  
Genew; HGNC:6763; MAD2L1.  
GK; Q13257; -;  
MIM; 601467; -;  
GO; GO:0005699; C:kinetochore; TAS.  
GO; GO:0007067; P:mitosis; TAS.  
GO; GO:0007093; P:mitotic checkpoint; TAS.  
InterPro; IPR003511; DNABind\_HORVA.  
Pfam; PF02301; HORVA; 1.  
PROSITE; PS00815; HORVA; 1.  
Cell cycle; Mitosis; Nuclear protein; 3D-structure.  
DOMAIN 14 197  
STRAND 11 12  
STRAND 15 15  
HELIIX 17 34  
TURN 35 35  
TURN 41 42  
TURN 44 48  
TURN 49 50  
TURN 51 55  
STRAND 59 73  
HELIIX 83 90  
TURN 91 94  
STRAND 95 103  
STRAND 106 106  
TURN 112 113  
HELIIX 122 141  
STRAND 149 158  
TURN 165 166  
STRAND 173 175  
TURN 179 181  
STRAND 185 193  
SEQUENCE 205 AA; B8DCBF0043836764 CRC64;  
Query Match 95.0%; Score 19; DB 1; Length 205;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
2 VAEF 5  
|||  
20 VAEF 23  
-----  
SULT 61  
21 MOUSE  
MD21 MOUSE STANDARD; PRT; 205 AA.  
Q921B5; Q921B5;  
16-OCT-2001 (Rel. 40, Created)  
16-OCT-2001 (Rel. 40, Last sequence update)  
10-OCT-2003 (Rel. 42, Last annotation update)  
Mitotic spindle assembly checkpoint protein MAD2A (MAD2-like 1).  
-----

GN MAD2L1 OR YAD2A.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryo;  
RA Jin D.-Y., Jeang K.-T.;  
RT "Identification of a novel component of the spindle assembly  
RT checkpoint in mammalian cells";  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129;  
RX MEDLINE=20348739; PubMed=10892650;  
RA Dobles M., Liberal V., Scott M.L., Benezra R., Sorger P.K.;  
RT "Chromosome missegregation and apoptosis in mice lacking the mitotic  
RT checkpoint protein Mad2";  
RL Cell 101:635-645 (2000).  
CC -1- FUNCTION: REQUIRED FOR THE EXECUTION OF THE MITOTIC CHECKPOINT  
CC WHICH MONITORS THE PROCESS OF KINETOCHORE-SPINDLE ATTACHMENT AND  
CC DELAYS THE ONSET OF ANAPHASE WHEN THIS PROCESS IS NOT COMPLETE. IT  
CC INHIBITS THE ACTIVITY OF THE ANAPHASE PROMOTING COMPLEX BY  
CC SEQUESTERING CDC20 UNTIL ALL CHROMOSOMES ARE ALIGNED AT THE  
CC METAPHASE PLATE (BY SIMILARITY).  
CC -1- SUBUNIT: INTERACTS WITH CDC20.  
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE MAD2 FAMILY.  
CC -1- SIMILARITY: Contains 1 HORVA domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U83902; AAD09238.1; -;  
CC EMBL; AF261919; AAF69525.1; -;  
CC HSSP; Q13257; 1DUU.  
CC MGD; MGI:1860374; Mad2L1.  
CC GO; GO:0000776; C:kinetochore; IDA.  
CC GO; GO:0005816; C:spindle pole body; IDA.  
CC GO; GO:0000070; P:mitotic chromosome segregation; IMP.  
CC GO; GO:0007094; P:mitotic spindle checkpoint; IMP.  
CC InterPro; IPR003511; DNABind\_HORVA.  
CC Pfam; PF02301; HORVA; 1.  
CC PROSITE; PS00815; HORVA; 1.  
CC Cell cycle; Mitosis; Nuclear protein.  
CC FT DOMAIN 14 197  
CC FT CONFLICT 157 157 T -> A (IN REF. 1).  
CC FT CONFLICT 178 178 C -> S (IN REF. 1).  
CC FT CONFLICT 201 201 T -> I (IN REF. 1).  
CC SEQUENCE 205 AA; 23598 MW; A9F3F28BC4C9738E CRC64;  
Query Match 95.0%; Score 19; DB 1; Length 205;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
2 VAEF 5  
|||  
20 VAEF 23  
-----  
RESULT 62  
YAC1 LEGPN  
ID YAC1 LEGPN STANDARD; PRT; 208 AA.  
AC P37033;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)

E Hypothetical 23.7 kDa protein in acc 5' region.  
S Legionella pneumophila.  
C Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;  
C Legionellaceae; Legionella.  
X NCBI\_TaxID=446;  
N [1]  
P SEQUENCE FROM N.A.  
C STRAIN=Philadelphia 1;  
X MEDLINE=93374864; PubMed=8366052;  
A Mengaud J.M., Horwitz M.A.;  
T "The major iron-containing protein of Legionella pneumophila is an  
T aconitase homologous with the human iron-responsive element-binding  
T protein.";  
T J. Bacteriol. 175:5666-5676 (1993).  
L -----  
C This SWISS-PROT entry is copyright. It is produced through a collaboration  
C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
C the European Bioinformatics Institute. There are no restrictions on its  
C use by non-profit institutions as long as its content is in no way  
C modified and this statement is not removed. Usage by and for commercial  
C entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
L -----  
C EMBL; L22081; AAA25294.1; -.  
R PIR; A48642; A48642.  
W Hypothetical protein.  
Q SEQUENCE 208 AA; 23714 MW; 1E5C75E63A20C800 CRC64;  
Query Match 95.0%; Score 19; DB 1; Length 208;  
Best Local Similarity 100.0%; Pred.No. 3.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 2 VAEF 5  
b 45 VAEF 48  
L -----  
E RESULT 63  
MUC CORDI  
D "HMO CORDI STANDARD; PRT; 215 AA.  
C P71119;  
T 15-DEC-1998 (Rel. 37, Created)  
T 15-MAR-2004 (Rel. 43, Last sequence update)  
T 15-MAR-2004 (Rel. 43, Last annotation update)  
E Heme oxygenase (EC 1.14.99.3).  
N HMO OR DIP1669.  
S Corynebacterium diphtheriae.  
C Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
C Corynebacterinaceae; Corynebacteriaceae; Corynebacterium.  
X NCBI\_TaxID=1717;  
X [1]  
P SEQUENCE FROM N.A.  
C STRAIN=C7;  
X MEDLINE=97158681; PubMed=9006041;  
A Schmitt M.P.;  
T "Utilization of host iron sources by Corynebacterium diphtheriae:  
T identification of a gene whose product is homologous to eukaryotic  
T heme oxygenases and is required for acquisition of iron from heme and  
T hemoglobin.";  
T J. Bacteriol. 179:838-845 (1997).  
L [2]  
P SEQUENCE FROM N.A.  
C STRAIN=Biotype Gravis / NCTC 13129;  
X MEDLINE=22955443; PubMed=14602910;  
A Cerdono-Tarraga A.M., Efstratiou A., Dover L.G., Holden M.T.G.,  
A Pallen M., Bentley S.D., Beira G.S., Churcher C., James K.D.,  
A De Zeyza A., Chillingworth T., Cronin A., Dowd L., Feltwell T.,  
A Hamlin N., Holroyd S., Jagels K., Moule S., Quail M.A.,  
A Rebbinowitsch E., Rutherford K.M., Thomson N.R., Unwin L.,  
A Whitehead S., Barrall B.G., Parkhill J.;  
T "The complete genome sequence and analysis of Corynebacterium  
T diphtheriae NCTC13129.";  
T Nucleic Acids Res. 31:6516-6523 (2003).  
L -----

CC -!- FUNCTION: Allows the bacteria to use the host heme as an iron  
CC source. Involved in the oxidation of heme and subsequent release  
CC of iron from the heme moiety.  
CC -!- CATALYTIC ACTIVITY: Heme + 3 AH(2) + O(2) = biliverdin + Fe(2+) +  
CC CO + 3 A + 3 H(2)O.  
CC [1]  
CC -!- SIMILARITY: Belongs to the heme oxygenase family.  
L -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
L -----  
CC EMBL; U73860; AAC44832.1; -.  
DR EMBL; EX248358; CAE50198.1; -.  
DR HSSP; P09601; 1Q08.  
DR InterPro; IPR002051; Heme\_oxygenase.  
DR Pfam; PF01126; Heme\_oxygenase; 1.  
DR PRINTS; PR00088; HEME\_OXYGENASE.  
DR PROSITE; PS00593; HEME\_OXYGENASE; 1.  
KW Heme; Oxidoreductase.  
FT METAL 20 20 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
FT CONFLICT 34 34 E -> K (IN REF. 1).  
FT CONFLICT 60 60 A -> V (IN REF. 1).  
FT CONFLICT 92 93 DG -> GS (IN REF. 1).  
FT CONFLICT 192 192 N -> H (IN REF. 1).  
SQ SEQUENCE 215 AA; 24116 MW; 60D9B8E2ED7ED456 CRC64;  
Query Match 95.0%; Score 19; DB 1; Length 215;  
Best Local Similarity 100.0%; Pred.No. 3.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 VAEF 5  
Db 39 VAEF 42  
L -----  
E RESULT 64  
COAT CMVIX STANDARD; PRT; 218 AA.  
AC Q66120;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Coat protein (Capsid protein) (CP).  
OS Cucumber mosaic virus (strain Ixora) (CMV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;  
OC Cucumovirus.  
OX NCBI\_TaxID=117114;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96005047; PubMed=7561763;  
RA McGarvey P.B., Tounignant M., Geletka L., Cellini F., Kaper J.M.;  
RT "The complete sequence of a cucumber mosaic virus from Ixora that is  
RT deficient in the replication of satellite RNAs.";  
RL J. Gen. Virol. 76:2257-2270 (1995).  
CC -!- SIMILARITY: Belongs to the cucumoviruses/bromoviruses coat protein  
CC family.  
L -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
L -----  
CC EMBL; U20219; AAC54619.1; -.  
DR PIR; D71392; D71392.  
DR InterPro; IPR000247; Cucumovirus\_coat.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
L -----

```

1 Pfam; PF00760; Cucumo coat; 1.
2 PRINTS; P00222; CUCUMOCOAT.
3 ProDom; PD001284; Cucumovirus_coat; 1.
4 Coat protein.
5 SEQUENCE 218 AA; 24185 MW; 130E82D17BD75224 CRC64;
Query Match 95.0%; Score 19; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2 VAEF 5
36 VAEF 99
SULT 65
P1_NPVOP STANDARD; PRT; 220 AA.
O10274;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Putative protein-tyrosine phosphatase 1 (SC 3.1.3.48).
PTP-1.
Orygia pseudotsugata multicapsid polyhedrosis virus (OPMPV).
Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.
NCBI_TaxID=164623;
[1]
SEQUENCE FROM N.A.
MEDLINE=97271300; PubMed=9126251;
Ahrens C.H.; Russell R.R.; Funk C.J.; Evans J.; Harwood S.;
Rohrmann G.F.;
"The sequence of the Orygia pseudotsugata multinucleocapsid nuclear
polyhedrosis virus genome."
Virology 229:381-395 (1997).
-!- FUNCTION: COULD BE INACTIVE AS THE ACTIVE SITE CYSTEINE IS
MODIFIED TO TRYPHOPHAN.
-!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
tyrosine + phosphate.
-!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
TYROSINE PHOSPHATASE FAMILY. CDC14 SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; U75930; AAC59009.1; -.
InterPro; IPR000387; TYR_phosphatase.
PROSITE; PS00383; TYR_PHOSPHATASE_1; FALSE_NEG.
PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
Hydrolase.
SITE 162 162 EQUIVALENT OF ACTIVE SITE CYS.
SEQUENCE 220 AA; 25234 MW; D3FC093F1953D425 CRC64;
Query Match 95.0%; Score 19; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2 VAEF 5
139 VAEF 142
SULT 66
R1_ARCFU STANDARD; PRT; 223 AA.
O29758;
10-OCT-2003 (Rel. 42, Created)

```

```

DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable exosome complex RNA-binding protein 1.
GN AF0492.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=96049343; PubMed=9389475;
RA Klenk H.-P.; Clayton R.A.; Tomb J.-F.; White O.; Nelson K.E.;
RA Ketchum K.A.; Dodson R.J.; Gwinn M.; Hickey E.K.; Peterson J.D.;
RA Richardson D.L.; Kerlavage A.R.; Graham D.E.; Kyrtides N.C.;
RA Fleischmann R.D.; Quackenbush J.; Lee N.H.; Sutton G.G.; Gill S.;
RA Kirkness E.F.; Dougherty B.A.; McKenney K.; Adams M.D.; Loftus B.;
RA Peterson S.; Reich C.I.; McNeil L.K.; Badger J.H.; Glöckle A.; Zhou L.;
RA Overbeek R.; Gocayne J.D.; Weidman J.F.; McDonald L.; Utterback T.;
RA Cotton M.D.; Spriggs T.; Artiach P.; Kaine B.P.; Sykes S.M.;
RA Sadow P.W.; D'Andrea K.P.; Bowman C.; Fujii C.; Garland S.A.;
RA Mason T.M.; Olsen G.J.; Fraser C.M.; Smith H.O.; Woese C.R.;
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 380:364-370 (1997).
CC -!- FUNCTION: Probably involved in degradation of a variety of RNA
species; could act a RNA-binding component of the exosome
(Potential).
CC -!- SUBUNIT: Component of the archaeal exosome multienzyme
ribonuclease complex (Potential).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Contains 1 KH domain.
CC -!- SIMILARITY: Contains 1 S1 motif domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; AE001070; AAB90745.1; -.
PIR; D69311; D69311.
TIGR; AF0492; -.
HAWAP; NF_00623; -.
InterPro; IPR004087; KH_dom.
InterPro; IPR004088; KH_type_1.
InterPro; IPR008994; Nucleic_acid_OB.
InterPro; IPR003029; S1.
Pfam; PF00013; KH; 1.
Pfam; PF00575; S1; 1.
SMART; SM00322; KH; 1.
SMART; SM00316; S1; 1.
DR PROSITE; PS00084; KH_TYPE_1; 1.
DR PROSITE; PS0126; S1; 1.
EXOSOME; RNA-binding; Complete proteome.
FT DOMAIN 58 127 S1 MOTIF.
FT DOMAIN 135 193 KH.
SQ SEQUENCE 223 AA; 24725 MW; A11B7F1A26972833 CRC64;
Query Match 95.0%; Score 19; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2 VAEF 5
209 VAEF 212
Oy
Db
RESULT 67
RAN_TETPY

```

D RAN TETPY STANDARD; PRT; 223 AA.  
 C P41514;  
 T 01-NOV-1995 (Rel. 32, Created)  
 T 01-NOV-1995 (Rel. 32, Last sequence update)  
 T 28-FEB-2003 (Rel. 41, Last annotation update)  
 E GTP-binding nuclear protein RAN/TC4.  
 S Tetrahymena pyriformis.  
 C Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;  
 X Tetrahymenina; Tetrahymena.  
 X NCBI\_TaxID=5908;  
 N [1]  
 P SEQUENCE FROM N.A.  
 C STRAIN=W;  
 X MEDLINE=94299154; PubMed=8026746;  
 A Nagata K.-I., Takemasa T., Alam S., Hattori T., Watanabe Y.,  
 T Nozawa Y.;  
 A "Cloning of cDNAs encoding a cell-cycle-regulatory GTP-binding  
 T low-M(r) (GBLM) protein, Ran/TC4, from micronucleated Tetrahymena  
 T thermophila and amiconucleated Tetrahymena pyriformis.";  
 L Gene 144:123-125(1994).  
 C -1- FUNCTION: GTP-binding protein involved in nucleocytoplasmic  
 C transport. Required for the import of protein into the nucleus and  
 C also for RNA export. Involved in chromatin condensation and  
 C control of cell cycle (By similarity).  
 C -1- SUBUNIT: Monomer.  
 C -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 C -1- SIMILARITY: Belongs to the small GTPase superfamily. Ran family.  
 C This SWISS-PROT entry is copyright. It is produced through a collaboration  
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 C the European Bioinformatics Institute. There are no restrictions on its  
 C use by non-profit institutions as long as its content is in no way  
 C modified and this statement is not removed. Usage by and for commercial  
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 C  
 C EMBL; D21825; BAA04849.1; --  
 R HSSP; P17080; 1A2K.  
 R InterPro; IPR002041; RAN.  
 R InterPro; IPR001806; Ras\_trnsfrmg.  
 R InterPro; IPR005225; Small\_GTP.  
 R Pfam; PF00071; ras; 1.  
 R PRINTS; PR00627; GTPRANTC4.  
 R SMART; SM00176; RAN; 1.  
 R TIGRfam; TIGR00231; small\_GTP; 1.  
 R PROSITE; PS01115; RAN; 1.  
 W GTP-binding; Nuclear protein; Protein transport.  
 T NP\_BIND 18 25  
 T NP\_BIND 66 70 GTP (By similarity).  
 T NP\_BIND 123 126 GTP (By similarity).  
 T D0MAIN 128 143 IBB (By similarity).  
 Q SEQUENCE 223 AA; 547ADADB5ABF0977 CRC64;  
 Query Match 95.0%; Score 19; DB 1; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 2 VAEF 5  
 b 9 VAEF 12  
 RESULT 68  
 ATD TETH  
 D -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 C -1- SIMILARITY: Belongs to the small GTPase superfamily. Ran family.  
 C This SWISS-PROT entry is copyright. It is produced through a collaboration  
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 C the European Bioinformatics Institute. There are no restrictions on its  
 C use by non-profit institutions as long as its content is in no way  
 C modified and this statement is not removed. Usage by and for commercial  
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 C  
 C EMBL; D21825; BAA04849.1; --  
 R HSSP; P17080; 1A2K.  
 R InterPro; IPR002041; RAN.  
 R InterPro; IPR001806; Ras\_trnsfrmg.  
 R InterPro; IPR005225; Small\_GTP.  
 R Pfam; PF00071; ras; 1.  
 R PRINTS; PR00627; GTPRANTC4.  
 R SMART; SM00176; RAN; 1.  
 R TIGRfam; TIGR00231; small\_GTP; 1.  
 R PROSITE; PS01115; RAN; 1.  
 W GTP-binding; Nuclear protein; Protein transport.  
 T NP\_BIND 18 25  
 T NP\_BIND 66 70 GTP (By similarity).  
 T NP\_BIND 123 126 GTP (By similarity).  
 T D0MAIN 128 143 IBB (By similarity).  
 Q SEQUENCE 223 AA; 547ADADB5ABF0977 CRC64;  
 Query Match 95.0%; Score 19; DB 1; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 2 VAEF 5  
 b 9 VAEF 12

OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
 OC Thermus.  
 OC NCBI\_TaxID=274;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HB8 / ATCC 27634;  
 RX MEDLINE=20250964; PubMed=10789522;  
 RA Yokoyama K., Ohkuma S., Taguchi H., Yasunaga T., Wakabayashi T.,  
 RA Yoshida M.;  
 RT "V-type H<sup>+</sup>-ATPase/synthase from a thermophilic eubacterium, Thermus  
 RT thermophilus. Subunit structure and operon.";  
 RL J. Biol. Chem. 275:13955-13961(2000).  
 CC -1- FUNCTION: Produces ATP from ADP in the presence of a proton  
 CC gradient across the membrane.  
 CC -1- CATALYTIC ACTIVITY: ATP + H<sup>2</sup>O + H<sup>+</sup> (In) = ADP + phosphate +  
 CC H<sup>+</sup> (Out).  
 CC -1- SIMILARITY: Belongs to the V-ATPase D subunit family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D63799; BAA33198.2; --  
 DR HAMAP; MF\_00271; --; 1.  
 DR InterPro; IPR002699; ATPsynt\_Dsub.  
 DR Pfam; PF01813; ATP-synt D; 1.  
 DR ProDom; PD004122; ATPsynt\_Dsub; 1.  
 DR TIGRfam; TIGR00309; V-ATPase\_subd; 1.  
 KW Hydrolase; ATP synthetase; Hydrogen ion transport.  
 SQ SEQUENCE 223 AA; 24677 MW; B666AE00DFE0B08B CRC64;  
 Query Match 95.0%; Score 19; DB 1; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 VAEF 5  
 Db 36 VAEF 39  
 RESULT 69  
 ID -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 AC RAN TETH STANDARD; PRT; 225 AA.  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE GTP-binding nuclear protein RAN/TC4.  
 OS Tetrahymena thermophila.  
 OS Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;  
 OC Tetrahymenina; Tetrahymena.  
 OC NCBI\_TaxID=5911;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B1868;  
 RX MEDLINE=94299154; PubMed=8026746;  
 RA Nagata K.-I., Takemasa T., Alam S., Hattori T., Watanabe Y.,  
 RA Nozawa Y.;  
 RT "Cloning of cDNAs encoding a cell-cycle-regulatory GTP-binding  
 RT low-M(r) (GBLM) protein, Ran/TC4, from micronucleated Tetrahymena  
 RT thermophila and amiconucleated Tetrahymena pyriformis.";  
 RL Gene 144:123-125(1994).  
 CC -1- FUNCTION: GTP-binding protein involved in nucleocytoplasmic  
 CC transport. Required for the import of protein into the nucleus and  
 CC also for RNA export. Involved in chromatin condensation and  
 CC control of cell cycle (By similarity).  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: Belongs to the small GTPase superfamily. Ran family.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; D17748; BAA04600.1; --  
 HSSP; P17080; 1A2K.  
 InterPro; IPR002041; RAN.  
 InterPro; IPR001806; Ras trnsfrmg.  
 InterPro; IPR005225; Small\_GTP.  
 Pfam; PF00071; ras; 1  
 PRINTS; PR00627; GTPRNTC4.  
 PRINTS; PR00449; RASTRNSFRMG.  
 SMART; SM00176; RAN; 1.  
 TIGRFAMs; TIGR00231; small\_GTP; 1.  
 PROSITE; PS01115; RAN; 1.  
 GTP-binding; Nuclear protein; Protein transport.  
 NP\_BIND 18 25 GTP (BY SIMILARITY).  
 NP\_BIND 66 70 GTP (BY SIMILARITY).  
 NP\_BIND 123 126 GTP (BY SIMILARITY).  
 DOMAIN 128 143 IBB (BY SIMILARITY).  
 SEQUENCE 225 AA; 25649 MW; 69F8B8409AF703F8 CRC64;

Query Match 95.0%; Score 19; DB 1; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 2 VAEF 5  
 ||||  
 9 VAEF 12

SUITE 70  
 32 NPVOP  
 Y132 NPVOP STANDARD; PRT; 228 AA.  
 P24080;  
 01-MAR-1992 (Rel. 21, Created)  
 01-MAR-1992 (Rel. 21, Last sequence update)  
 01-NOV-1997 (Rel. 35, Last annotation update)  
 Hypothetical 26.0 kDa protein in P934-EXO intergenic region (ORF 4).  
 Orygia pseudotsugata multicapsid polyhedrosis virus (OpNPV).  
 Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 Nucleopolyhedrovirus.  
 NCBI\_TaxID=164623;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=89293088; PubMed=2661722;  
 Gombart A.F., Blissard G.W., Rohrmann G.F.;  
 "Characterization of the genetic organization of the HindIII M region of the multicapsid nuclear polyhedrosis virus of Orygia pseudotsugata reveals major differences among baculoviruses.";  
 J. Gen. Virol. 70:1815-1828(1989).  
 [2]  
 SEQUENCE FROM N.A.  
 MEDLINE=97271300; PubMed=9126251;  
 Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S., Rohrmann G.F.;  
 "The sequence of the Orygia pseudotsugata multinucleocapsid nuclear polyhedrosis virus genome";  
 Virology 229:381-399(1997).  
 -!- SIMILARITY: TO CORRESPONDING ORF IN ACNPV.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

CC EMBL; D13796; BAA02952.1; --  
 DR EMBL; D13929; BAA03030.1; --  
 DR EMBL; U75930; AAC59129.1; --  
 DR PIR; D30857; D30857.  
 KW Hypothetical protein; Late protein.  
 SQ SEQUENCE 228 AA; 25962 MW; F0766AD21F60039D CRC64;  
 -----  
 Query Match 95.0%; Score 19; DB 1; Length 228;  
 Best Local Similarity 100.0%; Pred. No. 4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 2 VAEF 5  
 ||||  
 84 VAEF 87  
 -----  
 RESULT 71  
 NADD\_PSEPK STANDARD; PRT; 230 AA.  
 ID NADD\_PSEPK  
 AC Q88DL5;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Probable nicotinate-nucleotide adenylyltransferase (EC 2.7.7.18)  
 DE (Deamido-NAD(+)) pyrophosphorylase (Deamido-NAD(+)) diphosphorylase)  
 DE (Nicotinate mononucleotide adenylyltransferase) (NADN)  
 DE adenylyltransferase).  
 GN NADD OR PP4810.  
 OS Pseudomonas putida (strain KT2440).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 CC NCBI\_TaxID=160488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22423060; PubMed=12534463;  
 RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,  
 RA Martins Cos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,  
 RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,  
 RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,  
 RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzes A.,  
 RA Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,  
 RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,  
 RA Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuemmler B.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440.";  
 RL Environ. Microbiol. 4:799-808(2002).  
 CC -!- FUNCTION: Catalyzes the reversible adenylation of nicotinate mononucleotide (NADN) to nicotinic acid adenine dinucleotide (NAD) (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + nicotinate ribonucleotide = diphosphate + deamido-NAD(+).  
 CC -!- PATHWAY: NAD biosynthesis.  
 CC -!- SIMILARITY: Belongs to the nadd family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; AE016792; AAN70379.1; --  
 DR TIGR; PP4810; --  
 DR HAMAP; MF\_00244; --; 1.  
 DR InterPro; IPR004820; Cytidylyltransf.  
 DR Pfam; PF01467; CTP transf. 2; 1.  
 KW Pyridine nucleotide biosynthesis; Transferase; Nucleotidytransferase;  
 KW NAD; Complete proteome.  
 SQ SEQUENCE 230 AA; 25291 MW; 4CF04E8F70E48941 CRC64;

```
Query Match          95.0%; Score 19; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
b 42 VAEF 45

RESULT 72
R3_PORPU STANDARD; PRT; 230 AA.
C P51308;
I 01-OCT-1996 (Rel. 34, Created)
I 01-OCT-1996 (Rel. 34, Last sequence update)
I 28-FEB-2003 (Rel. 41, Last annotation update)
E Chloroplast 30S ribosomal protein S3.
V RPS3.
V Porphyrin purpurea.
V Chloroplast.
V Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
X NCBI_TaxID=2787;
Y SEQUENCE FROM N.A.
P STRAIN=Avonport;
C Reith M.E., Munnoland J.;
I "Complete nucleotide sequence of the Porphyra purpurea chloroplast
  genome.";
L Plant Mol. Biol. Rep. 13:333-335(1995).
C -1- SUBUNIT: Part of the 30S ribosomal subunit.
C -1- SUBCELLULAR LOCATION: Chloroplast.
C -1- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
C -1- SIMILARITY: Contains 1 KH type-2 domain.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announcement/
C or send an email to license@isb-sib.ch).
C -----
C EMBL; U38804; AAC08194.1; -
C FIR; S73229; S73229.
C HAMAP; MF 01309; -; 1.
C InterPro; IPR004087; KH dom.
C InterPro; IPR009019; KH_prok.
C InterPro; IPR004044; KH_TYPE_2.
C InterPro; IPR001351; Ribosomal_S3_C.
C InterPro; IPR008282; Ribosomal_S3_N.
C InterPro; IPR005704; S3_bact.
C Pfam; PF00013; KH; 1.
C Pfam; PF00189; Ribosomal_S3_C; 1.
C Pfam; PF00417; Ribosomal_S3_N; 1.
C SMART; SM00322; KH; 1.
C TIGRFSMS; TIGR01009; rpsc_bact; 1.
C PROSITE; PS50823; KH TYPE 2; 1.
C PROSITE; PS00548; RIBOSOMAL_S3; 1.
W Ribosomal protein; RNA-binding; rRNA-binding; Chloroplast.
T DOMAIN 39 109 KH TYPE-2.
Q SEQUENCE 230 AA; 25631 MW; 5489B93704D7F0ED CRC64;

Query Match          95.0%; Score 19; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
b 118 VAEF 121

RESULT 73
BIE_COREP
```

```
ID UBIE COREP STANDARD; PRT; 230 AA.
AC Q8FSE3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Menquinone biosynthesis methyltransferase ubiE (EC 2.1.1.-).
GN UBIE OR C80481.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RX MEDLINE=22723752; PubMed=12840036;
RA Nishio Y., Nakamura Y., Kawarabayashi Y., Usuda Y., Kimura E.,
RA Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,
RA Gojobori T.;
RT "Comparative complete genome sequence analysis of the amino acid
RT replacements responsible for the thermostability of Corynebacterium
RT efficiens.";
RL Genome Res. 13:1572-1579(2003).
CC -1- FUNCTION: Methyltransferase required for the conversion of
CC dimethylmenquinone (DMKH2) to menaquinone (MKH2) (By similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + demethylmenaquinol =
CC S-adenosyl-L-homocysteine + menaquinol.
CC -1- PATHWAY: Menaquinone biosynthesis; last step.
CC -1- SIMILARITY: Belongs to the ubiE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF005215; BAC17291.1; ALT_INIT.
CC HAMAP; MF 01813; -; 1.
CC InterPro; IPR001601; Methyltransf.
CC InterPro; IPR000051; SAM bind.
CC InterPro; IPR004033; UbiE/COQ5_Metr.
CC Pfam; PF01209; UbiE_methyltran; 1.
CC PROSITE; PS01183; UBIE_1; FALSE_NEG.
CC PROSITE; PS01184; UBIE_2; FALSE_NEG.
KW Menaquinone biosynthesis; Transferase; Methyltransferase;
KW Complete proteome.
SQ SEQUENCE 230 AA; 25366 MW; E10DCB602A1CD886 CRC64;

Query Match          95.0%; Score 19; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
Db 145 VAEF 148

RESULT 74
UBIE CORGL STANDARD; PRT; 230 AA.
AC Q8NT39;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Menaquinone biosynthesis methyltransferase ubiB (EC 2.1.1.-).
GN UBIE OR CGL0471.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=17118;
RN [1]
RP SEQUENCE FROM N.A.
```

STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
 Nakagawa S.;  
 "Complete genomic sequence of *Corynebacterium glutamicum* ATCC 13032.";  
 Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 -1- FUNCTION: Methyltransferase required for the conversion of  
 dimethylmenaquinone (DMKH2) to menaquinone (MKH2) (By similarity).  
 -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + demethylmenaquinol =  
 S-adenosyl-L-homocysteine + menaquinol.  
 -1- PATHWAY: Menaquinone biosynthesis; last step.  
 -1- SIMILARITY: Belongs to the ubiE family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; AP005275; BAB97864.1; -;  
 HAMAP; MF\_01813; -; 1;  
 InterPro; IPR001601; Methyltransf.  
 InterPro; IPR000051; SAM\_Bind.  
 InterPro; IPR004033; UbiE/COQ5\_Metrf.  
 Pfam; PF01209; UbiE\_methyltran; 1.  
 PROSITE; PS01183; UBI1; FALSE NEG.  
 PROSITE; PS01184; UBI2; FALSE NEG.  
 Menaquinone biosynthesis; Transferase; Methyltransferase;  
 Complete proteome.  
 SEQUENCE 230 AA; 25244 MW; A8548173B7901400 CRC64;  
 -----  
 Query Match 95.0%; Score 19; DB 1; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 -----  
 2 VAEF 5  
 ||||  
 145 VAEF 148  
 -----  
 SULT 75  
 YGIB\_ECOLI STANDARD; PRT; 234 AA.  
 P24195;  
 01-MAR-1992 (Rel. 21, Created)  
 01-MAR-1992 (Rel. 21, Last sequence update)  
 10-OCT-2003 (Rel. 42, Last annotation update)  
 Hypothetical protein ygiB.  
 YGIB OR B3037 OR C3783 OR SF3077 OR S3282.  
 Escherichia coli,  
 Escherichia coli O6, and  
 Shigella flexneri.  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.  
 NCBI\_TaxID=562, 217992, 523;  
 [1]  
 SEQUENCE FROM N.A.  
 SPECIES=E.coli;  
 MEDLINE=92223101; PubMed=1314093;  
 Yang T.-P., Depew R.E.;  
 "Nucleotide sequence of a region duplicated in *Escherichia coli* toc  
 mutants.";  
 Blochim. Biophys. Acta 1130:227-228(1992).  
 [2]  
 SEQUENCE FROM N.A.  
 SPECIES=E.coli; STRAIN=K12 / MG1655;  
 MEDLINE=97426617; PubMed=9278503;  
 Blattner P.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 Mau B., Shao Y.;  
 "The complete genome sequence of *Escherichia coli* K-12.";  
 Science 277:1453-1474(1997).

[3]  
 SEQUENCE FROM N.A.  
 SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 MEDLINE=22388234; PubMed=12471157;  
 Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 Mobley H.L.F., Donnenberg M.S., Blattner F.R.;  
 "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic *Escherichia coli*.";  
 Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 [4]  
 SEQUENCE FROM N.A.  
 SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
 MEDLINE=22272406; PubMed=12384590;  
 Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
 Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 Yu J.;  
 "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity  
 through comparison with genomes of *Escherichia coli* K12 and O157.";  
 Nucleic Acids Res. 30:4432-4441(2002).  
 [5]  
 SEQUENCE FROM N.A.  
 SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 MEDLINE=22590274; PubMed=12704152;  
 Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 Schwartz D.C., Blattner F.R.;  
 "Complete genome sequence and comparative genomics of *Shigella*  
 flexneri serotype 2a strain 2457T.";  
 Infect. Immun. 71:2775-2786(2003).  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; M77129; AAA71875.1; -;  
 EMBL; U28377; AAA69205.1; -;  
 DR EMBL; A3000385; AAC76073.1; -;  
 DR EMBL; A3016766; AAN82227.1; -;  
 DR EMBL; A3015319; AAN44555.1; ALT\_INIT.  
 DR EMBL; A3018988; AAP18368.1; -;  
 DR PIR; S22360; S22360.  
 DR EcoGene; EG1164; ygiB.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 234 AA; 24868 MW; 08D5DE931072C966 CRC64;  
 -----  
 Query Match 95.0%; Score 19; DB 1; Length 234;  
 Best Local Similarity 100.0%; Pred. No. 4.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 -----  
 QY 2 VAEF 5  
 ||||  
 Db 100 VAEF 103  
 -----  
 Search completed: May 24, 2004, 17:41:05  
 Job time : 9.57143 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

1 on: May 24, 2004, 17:33:08 ; Search time 35.7143 Seconds  
(without alignments)  
44.173 Million cell updates/sec

file: US-09-594-978A-3

effect score: 20

quence: 1 XVAEF 5

bring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

tal number of hits satisfying chosen parameters: 1017041

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

tabase :

- SPTREMBL 25:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phase:\*
  - 10: sp\_plant:\*
  - 11: sp\_prodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_virus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	19	95.0	20	6	Q9T518
2	19	95.0	22	2	Q9R5C0
3	19	95.0	30	5	P82214
4	19	95.0	33	16	O25518
5	19	95.0	35	17	Q82219
6	19	95.0	36	6	Q9TQ77
7	19	95.0	44	2	Q48420
8	19	95.0	45	16	Q830H9
9	19	95.0	48	16	O25478
10	19	95.0	51	16	Q70XLS
11	19	95.0	52	2	Q44941
12	19	95.0	54	16	Q8X3Y8
13	19	95.0	54	16	Q8X3Y8
14	19	95.0	61	2	Q8VSE8
15	19	95.0	61	2	Q9P566
16	19	95.0	68	12	Q99G29

17	19	95.0	68	16	Q82X12
18	19	95.0	69	10	Q7X7L1
19	19	95.0	69	16	Q8NPA7
20	19	95.0	71	13	Q8AUO2
21	19	95.0	71	13	Q8AUPO
22	19	95.0	72	2	P97252
23	19	95.0	74	16	Q97JUA
24	19	95.0	74	16	Q82MX0
25	19	95.0	77	6	Q86ZM8
26	19	95.0	77	16	Q88VS7
27	19	95.0	79	12	Q98582
28	19	95.0	80	2	Q9KI36
29	19	95.0	81	16	Q89WL6
30	19	95.0	84	16	Q9KAVA
31	19	95.0	84	16	Q821D2
32	19	95.0	84	16	Q7VPV6
33	19	95.0	85	3	Q43116
34	19	95.0	87	6	Q95149
35	19	95.0	87	9	Q7Y3Y9
36	19	95.0	88	2	Q98590
37	19	95.0	88	10	Q41185
38	19	95.0	89	2	Q9ZGY2
39	19	95.0	89	16	Q5CNX1
40	19	95.0	89	16	Q5CFD1
41	19	95.0	90	4	Q81ZX3
42	19	95.0	91	13	Q98TB2
43	19	95.0	94	2	Q8KY12
44	19	95.0	94	6	O46425
45	19	95.0	96	16	Q8UB01
46	19	95.0	96	16	Q89WX7
47	19	95.0	96	17	Q8U0G9
48	19	95.0	97	12	Q86535
49	19	95.0	97	16	Q8Y476
50	19	95.0	99	16	Q9HZ38
51	19	95.0	99	16	Q885M3
52	19	95.0	100	16	Q82UR2
53	19	95.0	101	2	Q8VS54
54	19	95.0	101	17	Q97YC2
55	19	95.0	102	16	Q81R27
56	19	95.0	103	4	Q7Z2N2
57	19	95.0	103	16	Q9F368
58	19	95.0	103	16	Q89MS9
59	19	95.0	104	9	Q856B5
60	19	95.0	105	5	Q8SVK0
61	19	95.0	105	5	Q856U1
62	19	95.0	105	17	Q96231
63	19	95.0	106	1	P71530
64	19	95.0	106	6	Q9EGJ5
65	19	95.0	106	10	O23682
66	19	95.0	107	3	Q9C1L1
67	19	95.0	107	5	Q81HV3
68	19	95.0	107	16	Q8NRX6
69	19	95.0	109	13	Q9PU82
70	19	95.0	109	16	O53333
71	19	95.0	109	16	Q7TX28
72	19	95.0	110	12	O11343
73	19	95.0	110	16	Q88SG3
74	19	95.0	111	17	Q8U296
75	19	95.0	112	5	Q86PI2
76	19	95.0	112	13	Q98GR4
77	19	95.0	112	13	Q98GR3
78	19	95.0	112	16	Q91315
79	19	95.0	113	2	P95437
80	19	95.0	113	2	Q89133
81	19	95.0	113	17	Q8TYL3
82	19	95.0	114	16	Q8U5K8
83	19	95.0	115	4	Q9Y5V1
84	19	95.0	115	17	Q9V203
85	19	95.0	116	11	Q8CD94
86	19	95.0	116	12	Q9DM67
87	19	95.0	118	4	Q8TCV1
88	19	95.0	118	10	Q7XLU4
89	19	95.0	119	2	Q9EXG6

90	19	95.0	119	5	Q95YG4	Q95YG4 caenorhabdi	163	19	95.0	140	13	Q8JTB5	Q8JTB5 agalychnis
91	19	95.0	119	9	Q853J0	Q853J0 mycobacteri	164	19	95.0	140	16	Q891C6	Q891C6 clostridium
92	19	95.0	119	16	Q92ZP9	Q92ZP9 rhizobium m	165	19	95.0	140	16	Q83LW2	Q83LW2 shigella fl
93	19	95.0	119	16	Q87R53	Q87R53 vibrio para	166	19	95.0	142	10	Q23641	Q23641 arabisopsis
94	19	95.0	120	2	Q9ALD4	Q9ALD4 uncultured	167	19	95.0	142	10	Q39368	Q39368 brassica ol
95	19	95.0	120	2	Q9ALH9	Q9ALH9 uncultured	168	19	95.0	142	10	Q41213	Q41213 brassica na
96	19	95.0	120	2	Q9ALB1	Q9ALB1 uncultured	169	19	95.0	142	10	Q42386	Q42386 streptomyc
97	19	95.0	120	2	Q31353	Q31353 bacillus ce	170	19	95.0	143	2	Q54526	Q54526 synchococc
98	19	95.0	120	2	Q9ALH7	Q9ALH7 uncultured	171	19	95.0	143	16	Q8DXS6	Q8DXS6 nitrosomona
99	19	95.0	120	2	Q9ALH2	Q9ALH2 uncultured	172	19	95.0	144	5	Q47428	Q47428 drosophila
100	19	95.0	120	2	Q9ALE0	Q9ALE0 uncultured	173	19	95.0	144	10	Q9FFG7	Q9FFG7 arabisopsis
101	19	95.0	120	2	Q9ALD3	Q9ALD3 uncultured	174	19	95.0	144	16	Q8ABF9	Q8ABF9 bacteroides
102	19	95.0	120	16	Q8A661	Q8A661 bacteroides	175	19	95.0	144	16	Q54808	Q54808 streptomyc
103	19	95.0	121	4	Q7Z2N5	Q7Z2N5 homo sapien	176	19	95.0	145	2	Q84127	Q84127 vibrio vuln
104	19	95.0	121	9	Q9B066	Q9B066 mycobacteri	177	19	95.0	145	2	Q84127	Q84127 vibrio vuln
105	19	95.0	122	10	Q7XJZ7	Q7XJZ7 oryza sativ	178	19	95.0	145	13	Q9DFE4	Q9DFE4 oncorhynch
106	19	95.0	122	16	Q9PHJ9	Q9PHJ9 xylella fas	179	19	95.0	145	16	Q55965	Q55965 synchocyst
107	19	95.0	122	16	Q9CJM8	Q9CJM8 pasteurella	180	19	95.0	146	5	Q8WTP1	Q8WTP1 ascaris suu
108	19	95.0	123	16	Q912J4	Q912J4 streptomyc	181	19	95.0	146	5	Q23842	Q23842 drosophila
109	19	95.0	123	16	Q912J4	Q912J4 streptomyc	182	19	95.0	146	10	Q94AG7	Q94AG7 arabisopsis
110	19	95.0	124	9	Q03376	Q03376 bacterioph	183	19	95.0	146	16	Q98JX5	Q98JX5 rhizobium l
111	19	95.0	124	16	Q8FQZ0	Q8FQZ0 corynebacte	184	19	95.0	146	16	Q98JX5	Q98JX5 rhizobium l
112	19	95.0	124	16	Q818K2	Q818K2 bacillus ce	185	19	95.0	146	16	Q8FEX6	Q8FEX6 escherichia
113	19	95.0	125	16	Q69643	Q69643 mycobacteri	186	19	95.0	147	5	Q23923	Q23923 drosophila
114	19	95.0	125	16	Q7TVZ0	Q7TVZ0 mycobacteri	187	19	95.0	147	5	Q24730	Q24730 drosophila
115	19	95.0	126	17	Q26336	Q26336 methanobact	188	19	95.0	147	16	Q69906	Q69906 streptomyc
116	19	95.0	126	2	Q933V9	Q933V9 anabaena so	189	19	95.0	147	17	Q9HPZ8	Q9HPZ8 halobacteri
117	19	95.0	126	6	Q9XSF8	Q9XSF8 bos taurus	190	19	95.0	148	11	Q9CRI8	Q9CRI8 mus musculu
118	19	95.0	126	12	Q84351	Q84351 simian viru	191	19	95.0	148	16	Q9JSS2	Q9JSS2 neisseria m
119	19	95.0	126	16	Q9PNZ4	Q9PNZ4 campylobact	192	19	95.0	149	2	Q81KN2	Q81KN2 uncultured
120	19	95.0	126	16	Q7WFX5	Q7WFX5 bordetella	193	19	95.0	149	16	Q88C75	Q88C75 rhizobium l
121	19	95.0	127	12	Q91BF3	Q91BF3 spodoptera	194	19	95.0	149	16	Q819M8	Q819M8 listeria mo
122	19	95.0	127	16	Q69714	Q69714 mycobacteri	195	19	95.0	149	17	Q8FRU2	Q8FRU2 corynebacte
123	19	95.0	127	16	Q928V1	Q928V1 listeria in	196	19	95.0	150	16	Q87LY8	Q87LY8 methanosarc
124	19	95.0	127	16	Q83ET6	Q83ET6 coxiella bu	197	19	95.0	150	16	Q92F86	Q92F86 listeria in
125	19	95.0	127	16	Q7TVS6	Q7TVS6 mycobacteri	198	19	95.0	151	11	Q8YAK5	Q8YAK5 listeria mo
126	19	95.0	128	2	Q938X8	Q938X8 acinetobact	199	19	95.0	151	16	Q9CVJ5	Q9CVJ5 mus musculu
127	19	95.0	128	2	Q83YU7	Q83YU7 acinetobact	200	19	95.0	151	16	Q9ABU9	Q9ABU9 caulobacter
128	19	95.0	128	10	Q9ZSM5	Q9ZSM5 dendrobium	201	19	95.0	152	9	Q88BU6	Q88BU6 bacterioph
129	19	95.0	128	16	Q91SK0	Q91SK0 salmoneilla	202	19	95.0	152	9	Q7Y5S0	Q7Y5S0 bacterioph
130	19	95.0	129	10	Q9SBB1	Q9SBB1 arabisopsis	203	19	95.0	153	10	Q9ASAL	Q9ASAL oryza sativ
131	19	95.0	129	12	Q84339	Q84339 simian viru	204	19	95.0	153	10	Q8817	Q8817 oryza sativ
132	19	95.0	130	9	Q8WSY1	Q8WSY1 bacterioph	205	19	95.0	153	16	Q8XR05	Q8XR05 ralestonia s
133	19	95.0	130	12	Q8V7C4	Q8V7C4 tt virus. o	206	19	95.0	153	16	Q88SK9	Q88SK9 lactobacill
134	19	95.0	131	2	Q84CY0	Q84CY0 gamma-prote	207	19	95.0	154	2	Q52066	Q52066 pantoaea agg
135	19	95.0	132	2	Q44307	Q44307 anabaena sp	208	19	95.0	154	3	Q86ZD3	Q86ZD3 fusarium sp
136	19	95.0	132	9	Q8SDF0	Q8SDF0 pseudomonas	209	19	95.0	155	16	Q7UWB1	Q7UWB1 rhodospirell
137	19	95.0	132	16	Q8G332	Q8G332 gamma-prote	210	19	95.0	156	16	Q9PMR4	Q9PMR4 campylobact
138	19	95.0	133	16	Q9RZR3	Q9RZR3 deinococcus	211	19	95.0	156	17	Q974D5	Q974D5 sulfolobus
139	19	95.0	133	16	Q8ZGZ5	Q8ZGZ5 yersinia pe	212	19	95.0	157	5	Q9GTJ0	Q9GTJ0 heterodera
140	19	95.0	133	16	Q8D9V9	Q8D9V9 vibrio vuln	213	19	95.0	157	16	Q8FA79	Q8FA79 escherichia
141	19	95.0	133	16	Q8U1Z0	Q8U1Z0 pyrococcus	214	19	95.0	157	16	Q8DMU7	Q8DMU7 synchococc
142	19	95.0	134	2	Q937K7	Q937K7 erwinia chr	215	19	95.0	159	12	Q84662	Q84662 paramenech
143	19	95.0	134	5	Q96898	Q96898 sacculina c	216	19	95.0	159	16	Q8FTW6	Q8FTW6 corynebacte
144	19	95.0	134	9	Q857W7	Q857W7 mycobacteri	217	19	95.0	160	2	Q93KL6	Q93KL6 uncultured
145	19	95.0	134	16	Q8Q2T7	Q8Q2T7 salmoneilla	218	19	95.0	161	2	Q99QI3	Q99QI3 shigella fl
146	19	95.0	134	16	Q8Z8C2	Q8Z8C2 salmoneilla	219	19	95.0	161	2	Q55215	Q55215 streptomyc
147	19	95.0	134	17	Q8QOF7	Q8QOF7 methanosarc	220	19	95.0	161	2	Q8VSI8	Q8VSI8 shigella fl
148	19	95.0	134	17	Q8QO93	Q8QO93 methanosarc	221	19	95.0	161	2	Q8VSI8	Q8VSI8 shigella fl
149	19	95.0	135	2	Q939V2	Q939V2 anabaena sp	222	19	95.0	161	2	Q8VSI8	Q8VSI8 shigella fl
150	19	95.0	135	16	Q968D5	Q968D5 rhizobium l	223	19	95.0	161	16	Q8Z2B1	Q8Z2B1 salmoneilla
151	19	95.0	135	16	Q968D5	Q968D5 rhizobium l	224	19	95.0	161	16	Q8Z2B1	Q8Z2B1 salmoneilla
152	19	95.0	136	16	Q968D5	Q968D5 rhizobium l	225	19	95.0	162	4	Q98S36	Q98S36 homo sapien
153	19	95.0	136	16	Q968D5	Q968D5 rhizobium l	226	19	95.0	162	11	Q8CG34	Q8CG34 mus sp. myb
154	19	95.0	137	16	Q968D5	Q968D5 rhizobium l	227	19	95.0	163	4	Q86W13	Q86W13 homo sapien
155	19	95.0	137	16	Q968D5	Q968D5 rhizobium l	228	19	95.0	163	13	Q7ZXF9	Q7ZXF9 mycobacteri
156	19	95.0	137	16	Q968D5	Q968D5 rhizobium l	229	19	95.0	163	16	Q07237	Q07237 xenopus lae
157	19	95.0	138	16	Q968D5	Q968D5 rhizobium l	230	19	95.0	163	16	Q8ZL98	Q8ZL98 salmoneilla
158	19	95.0	138	16	Q968D5	Q968D5 rhizobium l	231	19	95.0	163	16	Q8ZL98	Q8ZL98 salmoneilla
159	19	95.0	139	16	Q968D5	Q968D5 rhizobium l	232	19	95.0	163	16	Q8ZL98	Q8ZL98 salmoneilla
160	19	95.0	139	16	Q968D5	Q968D5 rhizobium l	233	19	95.0	163	16	Q8ZL98	Q8ZL98 salmoneilla
161	19	95.0	139	16	Q968D5	Q968D5 rhizobium l	234	19	95.0	164	16	Q8ZL98	Q8ZL98 salmoneilla
162	19	95.0	140	11	Q968D5	Q968D5 rhizobium l	235	19	95.0	164	16	Q8ZL98	Q8ZL98 salmoneilla

236	19	95.0	165	4	Q96ED1	Q96ed1 homo sapien	309	19	95.0	188	16	Q8XC46	Q8xc46 escherichia
237	19	95.0	185	5	Q8SPD1	Q8srd1 encephalito	310	19	95.0	188	17	Q8TNJ0	Q8tnj0 methanosarc
238	19	95.0	185	11	Q9DeS8	Q9d6s8 mus musculus	311	19	95.0	189	2	Q84CM0	Q84cm0 streptomyc
239	19	95.0	166	2	Q52661	Q52661 rhodobacter	312	19	95.0	189	5	Q20199	Q20199 caenorhabdi
240	19	95.0	166	10	Q8RZ96	Q8r296 oryza sativ	313	19	95.0	189	8	Q8WJ84	Q8wj84 eriosorus h
241	19	95.0	166	10	Q948R0	Q948r0 oryza sativ	314	19	95.0	189	10	Q07299	Q07299 zea mays (m
242	19	95.0	166	17	Q82VJ9	Q82vj9 pyrobaculum	315	19	95.0	189	16	Q8U549	Q8u549 agrobacteri
243	19	95.0	167	2	Q8GAC0	Q8gac0 arthrobacte	316	19	95.0	189	16	Q8FC12	Q8fc12 escherichia
244	19	95.0	168	2	Q47850	Q47850 pantoea agg	317	19	95.0	190	4	Q96F12	Q96f12 homo sapien
245	19	95.0	168	16	Q8ZFRV0	Q8zfrv0 yersinia pe	318	19	95.0	190	16	Q93RS6	Q93rs6 streptomyc
246	19	95.0	169	16	Q9X8N2	Q9x8n2 streptomyc	319	19	95.0	191	2	Q93KL3	Q93kl3 uncultured
247	19	95.0	169	16	Q7UKV3	Q7ukv3 rhodopirell	320	19	95.0	191	10	Q84LF2	Q84lf2 oryza sativ
248	19	95.0	169	16	Q7U5D8	Q7u5d8 synechococc	321	19	95.0	191	10	Q84MM8	Q84mm8 oryza sativ
249	19	95.0	170	16	Q93JC6	Q93jc6 streptomyc	322	19	95.0	191	16	Q8KKV1	Q8kkv1 vibrio chol
250	19	95.0	170	16	Q881F4	Q881f4 pseudomonas	323	19	95.0	191	16	Q8EAK2	Q8eak2 shewanella
251	19	95.0	171	10	Q84QL9	Q84ql9 oryza sativ	324	19	95.0	192	10	Q8L647	Q8l647 deschampsia
252	19	95.0	171	12	Q8QW15	Q8qw15 black queen	325	19	95.0	192	16	Q8UG64	Q8ug64 agrobacteri
253	19	95.0	172	10	Q9AWP5	Q9awp5 oryza sativ	326	19	95.0	192	16	Q9ADP7	Q9adp7 streptomyc
254	19	95.0	172	11	Q85179	Q85179 mus musculu	327	19	95.0	192	16	Q83D37	Q83d37 coxiella bu
255	19	95.0	172	11	Q8X164	Q8x164 mus musculu	328	19	95.0	193	2	P74983	P74983 yersinia en
256	19	95.0	172	11	Q55177	Q55177 mus musculu	329	19	95.0	193	16	Q9KY43	Q9ky43 streptomyc
257	19	95.0	172	17	Q9UZ36	Q9uz36 pyrococcus	330	19	95.0	193	16	Q82B77	Q82b77 streptomyc
258	19	95.0	172	17	Q8U1U0	Q8u1u0 pyrococcus	331	19	95.0	194	2	Q9EVD1	Q9evd1 erwinia amy
259	19	95.0	173	5	Q20526	Q20526 caenorhabdi	332	19	95.0	194	2	Q8GQ2	Q8gq2 serratia ma
260	19	95.0	173	5	Q8SWX9	Q8swx9 drosophila	333	19	95.0	194	2	Q8GEL2	Q8gx2 vibrio angu
261	19	95.0	174	11	Q8CIG2	Q8cig2 mus musculu	334	19	95.0	194	2	Q8X3V2	Q8x3v2 erwinia sp.
262	19	95.0	174	16	Q8ECL6	Q8ec12 mus musculu	335	19	95.0	194	16	Q9A7S1	Q9a7s1 caulobacter
263	19	95.0	174	16	Q33289	Q33289 mycobacteri	336	19	95.0	194	16	Q8UBS8	Q8ubs8 agrobacteri
264	19	95.0	174	16	Q7V6A9	Q7v6a9 prochloroco	337	19	95.0	194	10	Q9LMI9	Q9lmi9 arabidopsi
265	19	95.0	174	16	Q7YXZ2	Q7yxz2 mycobacteri	338	19	95.0	195	10	Q8MTJ4	Q8mtj4 bradyrhizob
266	19	95.0	175	11	Q925A9	Q925a9 rattus norv	339	19	95.0	195	16	Q8SG8	Q8sg8 lactobacill
267	19	95.0	175	16	Q8A2B2	Q8a2b2 bacteroides	340	19	95.0	195	16	Q93KM6	Q93km6 uncultured
268	19	95.0	175	16	Q82DE1	Q82de1 streptomyc	341	19	95.0	196	5	Q9GNR5	Q9gnr5 styloxychia
269	19	95.0	175	17	Q58465	Q58465 pyrococcus	342	19	95.0	196	5	Q93568	Q93568 caenorhabdi
270	19	95.0	176	16	Q7VD73	Q7vd73 prochloroco	343	19	95.0	196	5	Q9GNR7	Q9gnr7 styloxychia
271	19	95.0	176	17	Q976D7	Q976d7 sulfolobus	344	19	95.0	196	5	Q9GNR7	Q9gnr7 styloxychia
272	19	95.0	178	5	Q8MTA3	Q8mta3 drosophila	345	19	95.0	196	11	Q9CVX4	Q9cvx4 mus musculu
273	19	95.0	178	16	Q7VUJ4	Q7vj4 helicobacte	346	19	95.0	196	11	Q8FT59	Q8ft59 corynebacte
274	19	95.0	178	16	Q7UT14	Q7ut14 rhodopirell	347	19	95.0	196	16	Q8DGL7	Q8dgl7 synechococc
275	19	95.0	179	2	Q54157	Q54157 shigella fl	348	19	95.0	196	16	Q8Z2K7	Q8z2k7 salmonella
276	19	95.0	179	11	Q8R057	Q8r057 mus musculu	349	19	95.0	197	9	Q8H9Q1	Q8h9q1 uncultured
277	19	95.0	179	16	Q7VKQ4	Q7vkq4 haemophilus	350	19	95.0	197	9	Q8H9Q1	Q8h9q1 vibrio harv
278	19	95.0	179	16	Q7UD52	Q7ud52 shigella fl	351	19	95.0	197	17	Q8TZV4	Q8tzv4 pyrococcus
279	19	95.0	180	2	Q8KS07	Q8ksu7 gamma-prote	352	19	95.0	197	17	Q8KEL5	Q8krl5 rhodocyclu
280	19	95.0	180	10	Q8RXU1	Q8rxu1 arabidopsi	353	19	95.0	198	2	Q8KEL5	Q8krc6 rhizobium l
281	19	95.0	180	10	Q9M098	Q9m098 arabidopsi	354	19	95.0	198	16	Q9N0C6	Q9n0c6 rhizobium l
282	19	95.0	180	10	Q8LDP3	Q8ldx3 arabidopsi	355	19	95.0	198	16	Q50927	Q50927 nitrosomona
283	19	95.0	180	16	Q8ZPH9	Q8zph9 salmonella	356	19	95.0	199	4	Q9H3L5	Q9h3l5 homo sapien
284	19	95.0	180	16	Q8Z707	Q8z707 salmonella	357	19	95.0	199	4	Q8EX83	Q8ex83 homo sapien
285	19	95.0	180	16	Q83T30	Q83t30 salmonella	358	19	95.0	199	11	Q8BXC6	Q8bxc6 mus musculu
286	19	95.0	180	16	Q82EC1	Q82ec1 streptomyc	359	19	95.0	199	11	Q8BJI3	Q8bjj3 mus musculu
287	19	95.0	180	16	Q8IDE6	Q8ide6 bacillus ce	360	19	95.0	199	16	Q8XTM5	Q8xtm5 ralstonia s
288	19	95.0	180	16	Q05559	Q05559 mycobacteri	361	19	95.0	199	16	Q7U983	Q7u983 synechococc
289	19	95.0	180	16	Q8UAE1	Q8uae1 shigella fl	362	19	95.0	200	16	Q8ZMY2	Q8zmy2 salmonella
290	19	95.0	181	5	Q8MUA9	Q8mu9 ancylostoma	363	19	95.0	200	16	Q9L0A5	Q9l0a5 streptomyc
291	19	95.0	181	16	Q8Y8B5	Q8ysb5 anabaena sp	364	19	95.0	200	16	Q8ZQA6	Q8zqa6 salmonella
292	19	95.0	182	16	Q7V2G9	Q7v2g9 prochloroco	365	19	95.0	200	16	Q8Z7W8	Q8z7w8 salmonella
293	19	95.0	182	17	Q8TW2	Q8tw2 methanopyru	366	19	95.0	201	2	Q9X407	Q9x407 methylosulf
294	19	95.0	183	4	Q8TC09	Q8tcg9 homo sapien	367	19	95.0	201	10	Q48848	Q48848 arabidopsi
295	19	95.0	184	16	Q8X5C1	Q8x5c1 escherichia	368	19	95.0	201	16	Q98I27	Q98i27 rhizobium l
296	19	95.0	184	16	Q9ZLY9	Q9zly9 helicobacte	369	19	95.0	202	2	Q8RNS6	Q8rns6 lactobacill
297	19	95.0	184	16	Q87YT2	Q87yt2 pseudomonas	370	19	95.0	202	16	Q8Y8Y2	Q8y8y2 listeria mo
298	19	95.0	184	16	Q7VC97	Q7vc97 prochloroco	371	19	95.0	202	16	Q69982	Q69982 streptomyc
299	19	95.0	185	16	Q8UHU5	Q8uhus agrobacteri	372	19	95.0	202	16	P73372	P73372 synechocyst
300	19	95.0	185	16	Q9RZH9	Q9rzh9 deinococcus	373	19	95.0	203	2	Q9RBB8	Q9rbb8 arthrobacte
301	19	95.0	186	2	Q56745	Q56745 wolbachia s	374	19	95.0	203	10	Q949A7	Q949a7 oryza sativ
302	19	95.0	186	2	P72209	P72209 proteus mir	375	19	95.0	203	16	Q8PB95	Q8pb95 xanthomonas
303	19	95.0	187	2	Q93T67	Q93t67 riemerella	376	19	95.0	203	16	Q889Y8	Q889y8 clostridium
304	19	95.0	187	16	Q98FD5	Q98fd5 rhizobium l	377	19	95.0	203	16	Q88DQ5	Q88dq5 pseudomonas
305	19	95.0	187	16	Q92WB0	Q92wb0 rhizobium m	378	19	95.0	204	2	Q9R8K6	Q9rbk6 xanthomonas
306	19	95.0	188	5	Q8T317	Q8t317 drosophila	379	19	95.0	204	2	Q83XE6	Q83xe6 xanthomonas
307	19	95.0	188	16	Q83722	Q83722 traponema p	380	19	95.0	204	4	Q9NZ83	Q9nz83 homo sapien
308	19	95.0	188	16	Q8Z37	Q8zn37 yersinia pe	381	19	95.0	204	4	Q7Z2N7	Q7z2n7 homo sapien

382	19	95.0	204	16	Q92F26	Q92fz6 rickettsia	455	19	95.0	224	16	Q8NRP2	Q8nrp2 corynebacte
383	19	95.0	204	16	Q8PQC4	Q8pqc4 xanthomonas	456	19	95.0	225	16	Q8S6L5	Q8s6l5 rhizobium l
384	19	95.0	204	16	Q93J78	Q93j78 streptomyce	457	19	95.0	225	16	Q82SR9	Q82sr9 streptomyce
385	19	95.0	204	16	Q7UB59	Q7ub59 shigella fl	458	19	95.0	226	2	Q8RED7	Q8red7 arthrobacte
386	19	95.0	205	16	Q8X810	Q8x810 escherichia	459	19	95.0	226	2	Q848B9	Q848b9 streptomyce
387	19	95.0	205	16	Q8G5L8	Q8g5l8 bifidobacte	460	19	95.0	226	5	Q8WPD3	Q8wpd3 paracentrot
388	19	95.0	205	16	Q8F8I0	Q8f8i0 escherichia	461	19	95.0	226	11	Q8CSE1	Q8cse1 mus muscucu
389	19	95.0	205	16	Q8DVM4	Q8dvm4 streptococ	462	19	95.0	226	16	Q8YDM2	Q8ydm2 brucella me
390	19	95.0	205	16	Q8DVI4	Q8dvi4 shigella fl	463	19	95.0	226	16	Q8FUS4	Q8fus4 brucella su
391	19	95.0	206	2	Q8KWD3	Q8kwd3 agrobacteri	464	19	95.0	227	2	Q9KJZ2	Q9kjl2 pseudomonas
392	19	95.0	206	11	Q8COY5	Q8coy5 mus muscucu	465	19	95.0	227	17	Q8GRV7	Q8grv7 oryza sativ
393	19	95.0	206	16	Q8EC34	Q8ec34 shewanella	466	19	95.0	227	17	Q8ZVD4	Q8zvd4 pyrobaculum
394	19	95.0	207	10	Q8H4R2	Q8h4r2 oryza sativ	467	19	95.0	228	16	Q8ZGQ3	Q8zgg3 salmonella
395	19	95.0	208	16	Q8PD50	Q8pd50 xanthomonas	468	19	95.0	228	16	Q8RS24	Q8rs24 deinococcus
396	19	95.0	208	16	Q82AM2	Q82am2 streptomyce	469	19	95.0	228	16	Q8Z153	Q8z153 yersinia pe
397	19	95.0	209	2	Q8RBE6	Q8rbbe arthrobacte	470	19	95.0	228	16	Q8FSE2	Q8fse2 corynebacte
398	19	95.0	209	4	Q8BYW9	Q8byw9 homo sapien	471	19	95.0	228	16	Q8Q4X0	Q8q4x0 clostridium
399	19	95.0	209	16	Q8UPF5	Q8ufp5 agrobacteri	472	19	95.0	228	16	Q8T3T6	Q8t3t6 salmonella
400	19	95.0	210	16	Q8ZPD7	Q8zpd7 salmonella	473	19	95.0	229	10	Q9CAS5	Q9cas5 arabidopsis
401	19	95.0	210	16	Q8G4Z4	Q8g4z4 bifidobacte	474	19	95.0	229	16	Q66753	Q66753 aquifex aso
402	19	95.0	210	16	Q83L90	Q83l90 shigella fl	475	19	95.0	229	16	Q8CSD8	Q8csd8 staphylococ
403	19	95.0	211	10	Q7XA92	Q7xa92 fragaria an	476	19	95.0	229	16	Q832X6	Q832x6 enterococcu
404	19	95.0	211	16	Q8ZRS0	Q8zrs0 rhizobium m	477	19	95.0	230	2	Q9RBB0	Q9rbbo arthrobacte
405	19	95.0	211	16	Q8PR40	Q8pr40 xanthomonas	478	19	95.0	230	5	Q764D5	Q764d5 caenorhabdi
406	19	95.0	212	2	Q9RBE7	Q9rbbe7 arthrobacte	479	19	95.0	230	10	Q84RH7	Q84rh7 lycopersico
407	19	95.0	212	10	Q8RUH2	Q8ruh2 oryza sativ	480	19	95.0	230	16	Q8UHL0	Q8uhl0 agrobacteri
408	19	95.0	212	16	Q914X9	Q914x9 pseudomonas	481	19	95.0	230	16	Q8DOP9	Q8dof9 yersinia pe
409	19	95.0	213	10	Q8H3F6	Q8h3f6 oryza sativ	482	19	95.0	230	16	Q7WQY2	Q7wcy2 bordetella
410	19	95.0	213	16	Q8KPS9	Q8kps9 vibrio chol	483	19	95.0	230	16	Q7W204	Q7w204 bordetella
411	19	95.0	213	16	Q8YK80	Q8yk80 anabaena sp	484	19	95.0	230	16	Q7VUL1	Q7vul1 bordetella
412	19	95.0	213	16	Q8NTZ8	Q8ntz8 corynebacte	485	19	95.0	231	5	Q8MXX3	Q8mxx3 trypanosoma
413	19	95.0	213	16	Q87VGR	Q87vgr pseudomonas	486	19	95.0	231	16	Q81684	Q81684 drosophila
414	19	95.0	213	17	Q9HR09	Q9hr09 halobacteri	487	19	95.0	231	16	Q9HXR1	Q9hxr1 pseudomonas
415	19	95.0	214	2	Q54380	Q54380 lactococcus	488	19	95.0	231	16	Q982T2	Q982t2 rhizobium l
416	19	95.0	214	2	Q8KP00	Q8kp00 gamma-prote	489	19	95.0	231	16	Q92E94	Q92e94 listeria in
417	19	95.0	214	2	Q9RC94	Q9rc94 bacillus ha	490	19	95.0	231	16	Q8YX7	Q8yx7 brucella me
418	19	95.0	215	2	Q9Z480	Q9z480 agrobacteri	491	19	95.0	231	16	Q819H1	Q819h1 listeria mo
419	19	95.0	215	10	Q22136	Q22136 arabidopsis	492	19	95.0	231	16	Q8XUT7	Q8xut7 ralstonia s
420	19	95.0	215	16	Q98DH0	Q98dh0 rhizobium l	493	19	95.0	231	16	Q8FOL3	Q8fql3 corynebacte
421	19	95.0	216	16	Q8FQ35	Q8fq35 xanthomonas	494	19	95.0	232	2	Q9RBD1	Q9rbdl arthrobacte
422	19	95.0	216	2	Q9RBD2	Q9rbd2 arthrobacte	495	19	95.0	232	16	Q9A6A4	Q9asa4 caulobacteri
423	19	95.0	216	14	P97979	P97979 unidentified	496	19	95.0	232	16	P96241	P96241 mycobacteri
424	19	95.0	216	16	Q92Q81	Q92q81 rhizobium m	497	19	95.0	232	16	P96241	P96241 mycobacteri
425	19	95.0	216	16	Q8NMA0	Q8nma0 pseudomonas	498	19	95.0	232	17	Q8PVL0	Q8pvl0 methanosarc
426	19	95.0	217	16	Q8GCA2	Q8gca2 pseudomonas	499	19	95.0	234	2	Q9RBC9	Q9rbcr9 arthrobacte
427	19	95.0	217	17	Q8UIA8	Q8uia8 pyrococcus	500	19	95.0	234	2	Q9N327	Q9n327 caenorhabdi
428	19	95.0	218	2	Q9RKA4	Q9rka4 arthrobacte	501	19	95.0	234	10	Q9FQC4	Q9fqc4 zea mays (m
429	19	95.0	218	10	Q40630	Q40630 oryza sativ	502	19	95.0	234	10	Q94K12	Q94kl2 arabidopsis
430	19	95.0	218	10	Q91ZY5	Q91zy5 arabidopsis	503	19	95.0	234	10	Q9XIF8	Q9xif8 arabidopsis
431	19	95.0	218	16	Q91470	Q91470 pseudomonas	504	19	95.0	234	16	Q8YSQ0	Q8ysq0 anabaena sp
432	19	95.0	218	16	Q8ZES1	Q8zes1 yersinia pe	505	19	95.0	234	16	Q8XBP4	Q8xbp4 escherichia
433	19	95.0	218	16	Q8NBB6	Q8nbb6 pseudomonas	506	19	95.0	234	16	Q8LD13	Q8ld13 bacillus ce
434	19	95.0	219	16	Q92E71	Q92e71 listeria in	507	19	95.0	235	2	Q9AKL9	Q9akl9 rickettsia
435	19	95.0	219	16	Q69984	Q69984 streptomyce	508	19	95.0	235	2	Q9AKB8	Q9akb8 rickettsia
436	19	95.0	219	16	Q8AKES	Q8akes streptomyce	509	19	95.0	235	2	Q9AKG3	Q9akg3 rickettsia
437	19	95.0	219	16	Q7UGG1	Q7uggl rhodospirell	510	19	95.0	235	2	Q849C9	Q849c9 streptomyce
438	19	95.0	219	17	Q26842	Q26842 methanobact	511	19	95.0	235	2	Q8W6X8	Q8w6x8 cyanophaga
439	19	95.0	220	17	Q9HQA1	Q9hqa1 halobacteri	512	19	95.0	235	16	Q92XB1	Q92xb1 rhizobium m
440	19	95.0	221	4	Q96GF5	Q96gf5 homo sapien	513	19	95.0	235	16	Q92LJ4	Q92lj4 rickettsia
441	19	95.0	221	16	Q9KAB1	Q9kab1 bacillus ha	514	19	95.0	235	16	Q8XJZ8	Q8xjz8 ralstonia s
442	19	95.0	221	16	Q93J74	Q93j74 streptomyce	515	19	95.0	236	2	Q9JPA0	Q9jpa0 rhodococcus
443	19	95.0	221	16	Q9RDX2	Q9rdx2 streptomyce	516	19	95.0	236	10	Q9LVA8	Q9lva8 arabidopsis
444	19	95.0	222	2	Q9RBD9	Q9rbd9 arthrobacte	517	19	95.0	236	10	Q9SK05	Q9sk05 arabidopsis
445	19	95.0	222	16	Q9F9P8	Q9f9ps leptospira	518	19	95.0	236	16	Q8VBG5	Q8vbg5 agrobacteri
446	19	95.0	223	5	Q8SUL0	Q8sul0 encephalito	519	19	95.0	237	2	Q9RBC0	Q9rbco arthrobacte
447	19	95.0	223	10	Q94B57	Q94b57 arabidopsis	520	19	95.0	237	2	Q84B84	Q84b84 vibrio fisc
448	19	95.0	223	16	Q9ZBP8	Q9zbp8 streptomyce	521	19	95.0	237	16	Q9KFS9	Q9kfs9 bacillus ha
449	19	95.0	223	16	Q883K9	Q883k9 pseudomonas	522	19	95.0	237	16	Q98N57	Q98n57 rhizobium l
450	19	95.0	223	16	Q8XGZ1	Q8xgz1 salmonella	523	19	95.0	237	16	Q97IY4	Q97iy4 clostridium
451	19	95.0	224	3	Q03362	Q03362 saccharomyc	524	19	95.0	237	16	Q8EIK7	Q8eik7 oceanobacil
452	19	95.0	224	4	Q92799	Q92799 homo sapien	525	19	95.0	237	17	Q8TMA2	Q8tma2 methanosarc
453	19	95.0	224	11	Q99MG0	Q99mg0 rattus norv	526	19	95.0	238	2	Q9RBP4	Q9rbp4 rhodococcus
454	19	95.0	224	16	Q8Y8J7	Q8y8j7 listeria mo	527	19	95.0	238	16	Q9CIV5	Q9civ5 lactococcus

528	19	95.0	238	16	Q9L246	Q9L246 streptomyc	601	19	95.0	253	16	Q8X4N5	Q8X4N5 escherichia
529	19	95.0	239	3	Q9UVW7	Q9UVW7 candida alb	602	19	95.0	253	16	Q9AK07	Q9AK07 streptomyc
530	19	95.0	239	16	Q836K1	Q836K1 enterococcu	603	19	95.0	253	16	Q8FE98	Q8FE98 escherichia
531	19	95.0	239	16	Q82UL5	Q82UL5 nitrosomona	604	19	95.0	253	16	Q83QB3	Q83QB3 shigella fl
532	19	95.0	239	17	Q82UL5	Q82UL5 nitrosomona	605	19	95.0	253	16	Q83QB3	Q83QB3 shigella fl
533	19	95.0	240	17	Q30144	Q30144 archaeoglob	606	19	95.0	254	2	Q9VB95	Q9VB95 aeropyrum p
534	19	95.0	240	17	Q7WZF7	Q7WZF7 streptomyc	607	19	95.0	254	2	Q9REK0	Q9REK0 arthrobacte
535	19	95.0	241	10	Q9SVW3	Q9SVW3 arbidopsi	608	19	95.0	254	8	Q9RW7	Q9RW7 guillardia
536	19	95.0	241	10	Q7XR82	Q7XR82 oryza sativ	609	19	95.0	254	10	Q9A55	Q9A55 arbidopsi
537	19	95.0	241	16	Q7WJN5	Q7WJN5 bordetella	610	19	95.0	254	10	Q22536	Q22536 oryza sativ
538	19	95.0	241	16	Q7WA18	Q7WA18 bordetella	611	19	95.0	254	10	Q9ATM0	Q9ATM0 zea mays (m
539	19	95.0	242	5	Q81683	Q81683 drosophila	612	19	95.0	254	16	Q8ZGH5	Q8ZGH5 yersinia pe
540	19	95.0	242	5	Q81682	Q81682 drosophila	613	19	95.0	255	8	Q9RBE0	Q9RBE0 arthrobacte
541	19	95.0	242	16	Q8Y0T0	Q8Y0T0 ralstonia s	614	19	95.0	255	8	Q950Y2	Q950Y2 tetrachyena
542	19	95.0	242	16	Q8DH96	Q8DH96 synechococc	615	19	95.0	255	16	Q9JJK6	Q9JJK6 bradyrhizob
543	19	95.0	242	16	Q8CYH8	Q8CYH8 streptococc	616	19	95.0	256	2	Q9RBD5	Q9RBD5 arthrobacte
544	19	95.0	243	2	Q9AN57	Q9AN57 bradyrhizob	617	19	95.0	256	8	Q9T7M4	Q9T7M4 tetrachyena
545	19	95.0	243	16	Q8G0Y5	Q8G0Y5 brucella su	618	19	95.0	256	10	Q94HB4	Q94HB4 oryza sativ
546	19	95.0	244	6	Q8HXL4	Q8HXL4 sus scrofa	619	19	95.0	256	10	Q65325	Q65325 arbidopsi
547	19	95.0	244	16	Q8DSN5	Q8DSN5 streptococc	620	19	95.0	256	10	Q7XFL7	Q7XFL7 oryza sativ
548	19	95.0	244	16	Q7QX99	Q7QX99 rhodopirell	621	19	95.0	256	10	Q9WN21	Q9WN21 frangipani
549	19	95.0	245	10	Q940J5	Q940J5 arbidopsi	622	19	95.0	256	12	Q9WN21	Q9WN21 frangipani
550	19	95.0	245	10	Q7XK61	Q7XK61 oryza sativ	623	19	95.0	257	10	Q8GUK5	Q8GUK5 arbidopsi
551	19	95.0	246	6	Q28535	Q28535 mustela vis	624	19	95.0	257	11	Q9CP85	Q9CP85 mus musculu
552	19	95.0	246	16	Q8PT51	Q8PT51 xanthomonas	625	19	95.0	257	16	Q92L02	Q92L02 rhizobium m
553	19	95.0	246	16	Q8P6V1	Q8P6V1 xanthomonas	626	19	95.0	257	16	Q8NLT7	Q8NLT7 corynebacte
554	19	95.0	246	16	Q9RDB0	Q9RDB0 streptomyc	627	19	95.0	257	16	Q9X7Q4	Q9X7Q4 streptomyc
555	19	95.0	246	16	Q8FM47	Q8FM47 corynebacte	628	19	95.0	258	2	Q9RBB1	Q9RBB1 arthrobacte
556	19	95.0	246	16	Q89MY3	Q89MY3 bradyrhizob	629	19	95.0	258	13	Q78ZS8	Q78ZS8 xenopus lae
557	19	95.0	246	16	Q82AC3	Q82AC3 streptomyc	630	19	95.0	258	16	Q9XZN9	Q9XZN9 pseudomonas
558	19	95.0	247	2	Q48830	Q48830 lactobacill	631	19	95.0	258	16	Q9X8L2	Q9X8L2 streptomyc
559	19	95.0	247	16	Q88ZC3	Q88ZC3 lactobacill	632	19	95.0	258	16	Q7WCV7	Q7WCV7 bordetella
560	19	95.0	248	5	Q97008	Q97008 leishmania	633	19	95.0	258	16	Q7W5C3	Q7W5C3 bordetella
561	19	95.0	248	10	Q949B9	Q949B9 oryza sativ	634	19	95.0	258	16	Q7VZN8	Q7VZN8 bordetella
562	19	95.0	248	10	Q93935	Q93935 helianthus	635	19	95.0	258	17	Q9HLA1	Q9HLA1 thermoplasma
563	19	95.0	248	10	Q81216	Q81216 zea mays (m	636	19	95.0	259	2	Q938F2	Q938F2 eubacterium
564	19	95.0	248	10	Q84RL6	Q84RL6 zea mays (m	637	19	95.0	259	2	Q8VMN6	Q8VMN6 pseudomonas
565	19	95.0	248	10	Q7XA61	Q7XA61 oryza sativ	638	19	95.0	259	13	Q7ZT19	Q7ZT19 xenopus lae
566	19	95.0	248	10	Q82G59	Q82G59 yersinia pe	639	19	95.0	259	16	P74082	P74082 synechocyst
567	19	95.0	248	16	Q8A971	Q8A971 bacteroides	640	19	95.0	259	16	Q9PJ15	Q9PJ15 campylobact
568	19	95.0	248	16	Q8A971	Q8A971 bacteroides	641	19	95.0	259	16	Q98K61	Q98K61 rhizobium l
569	19	95.0	248	17	Q9YCF6	Q9YCF6 aeropyrum p	642	19	95.0	259	16	Q7U9R9	Q7U9R9 synechococc
570	19	95.0	249	5	Q62042	Q62042 caenorhabdi	643	19	95.0	259	17	Q8ZT71	Q8ZT71 pyrobaculum
571	19	95.0	249	10	Q9LXK3	Q9LXK3 hordeum vul	644	19	95.0	260	2	Q9RBA7	Q9RBA7 arthrobacte
572	19	95.0	249	10	Q9ATL9	Q9ATL9 zea mays (m	645	19	95.0	260	2	Q8G976	Q8G976 pseudomonas
573	19	95.0	249	10	Q7XU33	Q7XU33 oryza sativ	646	19	95.0	260	2	Q849Q8	Q849Q8 pseudomonas
574	19	95.0	249	12	Q919L4	Q919L4 culx nigri	647	19	95.0	260	16	Q8YVJ0	Q8YVJ0 anabaena sp
575	19	95.0	249	16	Q9PKD5	Q9PKD5 chlamydia m	648	19	95.0	260	16	Q8FS45	Q8FS45 corynebacte
576	19	95.0	249	16	Q8P449	Q8P449 xanthomonas	649	19	95.0	260	16	Q81EA0	Q81EA0 bacillus ce
577	19	95.0	249	16	Q93N41	Q93N41 leptospira	650	19	95.0	260	16	Q7WN67	Q7WN67 bordetella
578	19	95.0	249	16	Q82M75	Q82M75 streptomyc	651	19	95.0	260	16	Q7W0U7	Q7W0U7 bordetella
579	19	95.0	249	17	Q29512	Q29512 archaeoglob	652	19	95.0	260	16	Q7WV38	Q7WV38 bordetella
580	19	95.0	250	10	Q9ATL8	Q9ATL8 zea mays (m	653	19	95.0	261	2	Q66247	Q66247 actinobacil
581	19	95.0	250	10	Q84JF6	Q84JF6 arbidopsi	654	19	95.0	261	2	Q9XDR0	Q9XDR0 actinobacil
582	19	95.0	250	11	Q8C4H6	Q8C4H6 mus musculu	655	19	95.0	261	3	Q9P8F1	Q9P8F1 glomus moss
583	19	95.0	250	16	Q25228	Q25228 helicobacte	656	19	95.0	261	3	Q9P8E9	Q9P8E9 glomus moss
584	19	95.0	251	2	Q93A69	Q93A69 gamma-prote	657	19	95.0	261	6	Q95LN6	Q95LN6 macaca fasc
585	19	95.0	251	5	Q8SVK8	Q8SVK8 encephalito	658	19	95.0	261	11	Q8K2L3	Q8K2L3 mus musculu
586	19	95.0	251	11	Q9D4Z4	Q9D4Z4 mus musculu	659	19	95.0	261	11	Q9DCE1	Q9DCE1 mus musculu
587	19	95.0	251	17	Q23540	Q23540 escherichia	660	19	95.0	261	11	Q9JMJ6	Q9JMJ6 mus musculu
588	19	95.0	252	9	Q8HA20	Q8HA20 bacterioph	661	19	95.0	261	11	Q8BH40	Q8BH40 mus musculu
589	19	95.0	252	10	Q94CS9	Q94CS9 oryza sativ	662	19	95.0	261	11	Q9X8W5	Q9X8W5 bacillus ha
590	19	95.0	252	10	Q7XA62	Q7XA62 oryza sativ	663	19	95.0	261	16	Q987X7	Q987X7 rhizobium l
591	19	95.0	252	10	Q7XA62	Q7XA62 oryza sativ	664	19	95.0	261	16	Q8UCAL	Q8UCAL agrobacteri
592	19	95.0	252	16	Q9A534	Q9A534 caulobacter	665	19	95.0	261	16	Q8U893	Q8U893 pseudomonas
593	19	95.0	252	16	Q88U86	Q88U86 lactobacill	666	19	95.0	261	16	Q87YH4	Q87YH4 agrobacteri
594	19	95.0	252	16	Q7UT61	Q7UT61 rhodopirell	667	19	95.0	261	16	Q7WPP1	Q7WPP1 bordetella
595	19	95.0	252	17	Q97VY6	Q97VY6 sulfolobus	668	19	95.0	261	16	Q7WBP9	Q7WBP9 bordetella
596	19	95.0	252	17	Q97VY6	Q97VY6 sulfolobus	669	19	95.0	262	3	Q8RSJ2	Q8RSJ2 gamma-prote
597	19	95.0	253	2	Q8KH15	Q8KH15 pectobacter	670	19	95.0	262	3	Q9P8E8	Q9P8E8 glomus moss
598	19	95.0	253	5	Q95U55	Q95U55 drosophila	671	19	95.0	262	16	Q9WZP2	Q9WZP2 thermotoga
599	19	95.0	253	10	Q7XD92	Q7XD92 oryza sativ	672	19	95.0	262	16	Q8EDY1	Q8EDY1 shewanella
600	19	95.0	253	16	Q8ZFH9	Q8ZFH9 yersinia pe	673	19	95.0	262	17	Q8U0A9	Q8U0A9 pyrococcus



820	19	95.0	291	16	Q97LH6	Q97LH6 clostridium	893	19	95.0	306	13	Q7ZX07	Q7zx07 xenopus lae
821	19	95.0	291	16	Q84633	Q84633 chlamydia t	894	19	95.0	306	16	Q8U746	Q8u746 agrobacteri
822	19	95.0	292	2	Q33766	Q33766 spingomona	895	19	95.0	306	16	Q8PCN2	Q8pgn2 xanthomonas
823	19	95.0	292	5	Q16616	Q16616 caenorhabdi	896	19	95.0	306	17	Q8ZC4	Q8zcc4 pyrobaculum
824	19	95.0	292	9	Q7Y2C5	Q7y2c5 bacterioph	897	19	95.0	307	5	Q18327	Q18327 caenorhabdi
825	19	95.0	293	5	Q9GN9	Q9gny9 leishmania	898	19	95.0	307	5	Q8SIV9	Q8stv9 encephalito
826	19	95.0	293	16	Q83315	Q83315 treponema p	899	19	95.0	307	10	Q9LNR7	Q9lnr7 arabidopsis
827	19	95.0	293	16	Q88LZ3	Q88l23 pseudomonas	900	19	95.0	307	11	Q8VGH1	Q8vgh1 mus musculu
828	19	95.0	293	16	Q826J6	Q826j6 streptomyc	901	19	95.0	307	16	Q82PD4	Q82pd4 streptomyce
829	19	95.0	294	2	Q9RG4	Q9rgk4 bacteroides	902	19	95.0	308	11	Q8VFG8	Q8vfg8 mus musculu
830	19	95.0	294	4	Q9NRW3	Q9nrw3 homo sapien	903	19	95.0	308	16	Q9CH02	Q9ch02 lactococcus
831	19	95.0	294	10	Q8SILP1	Q8silp1 oryza sativ	904	19	95.0	308	16	Q8NL14	Q8nl14 corynebacte
832	19	95.0	295	2	Q9XDK6	Q9xdk6 bacteroides	905	19	95.0	308	16	Q9V947	Q9v947 prochloroco
833	19	95.0	295	9	Q854V0	Q854v0 mycobacteri	906	19	95.0	309	10	Q7R000	Q7rc00 arabidopsis
834	19	95.0	295	10	Q94DE7	Q94de7 oryza sativ	907	19	95.0	309	11	Q8VGU8	Q8vgu8 mus musculu
835	19	95.0	295	16	Q9PGP2	Q9pgp2 xyella fas	908	19	95.0	309	11	Q7TRC3	Q7trc3 mus musculu
836	19	95.0	295	16	Q9KY64	Q9kv64 vibrio chol	909	19	95.0	309	16	Q98RE1	Q98re1 mycoplasma
837	19	95.0	295	16	Q8KXN8	Q8kx8 streptococc	910	19	95.0	309	17	Q8PYV5	Q8pyv5 methanosarc
838	19	95.0	295	16	Q8AAK1	Q8aak1 bacteroides	911	19	95.0	310	4	Q9YSE6	Q9yse6 homo sapien
839	19	95.0	295	16	Q87ET9	Q87et9 xyella fas	912	19	95.0	310	10	Q9ZSC1	Q9zsc1 lycopersico
840	19	95.0	296	2	Q9LC87	Q9lc87 nocardioides	913	19	95.0	310	10	Q9SPD6	Q9spd6 arabidopsis
841	19	95.0	296	17	Q8TU61	Q8tu61 methanosarc	914	19	95.0	310	10	Q84Z99	Q84z99 oryza sativ
842	19	95.0	296	17	Q8PWS7	Q8pws7 methanosarc	915	19	95.0	310	11	Q8K2C6	Q8k2c6 mus musculu
843	19	95.0	297	2	Q33LQ8	Q33lq8 klebsiella	916	19	95.0	310	12	Q9BMO9	Q9bmq9 amastrea moo
844	19	95.0	297	10	Q82163	Q82163 arabidopsis	917	19	95.0	310	16	Q9CK67	Q9ck67 pasteurella
845	19	95.0	297	16	Q8Y53	Q8yy53 anabaena sp	918	19	95.0	310	16	Q8E877	Q8e877 pseudomonas
846	19	95.0	297	16	Q8XG56	Q8xg56 anabaena sp	919	19	95.0	311	10	Q8YF83	Q8yf83 pseudomonas
847	19	95.0	297	16	Q8XG56	Q8xgb0 anabaena sp	920	19	95.0	311	10	Q8YF83	Q8yf83 pseudomonas
848	19	95.0	297	16	Q8R977	Q8r977 thermoanaer	921	19	95.0	311	11	Q8VF62	Q8vf62 mus musculu
849	19	95.0	298	13	Q7ZV21	Q7zv21 brachydania	922	19	95.0	311	16	Q98N28	Q98n28 rhizobium l
850	19	95.0	298	16	Q8U7A1	Q8u7a1 agrobacteri	923	19	95.0	311	16	Q8NU00	Q8nu00 corynebacte
851	19	95.0	298	16	Q8ELP1	Q8elp1 oceanobacill	924	19	95.0	311	16	Q8FU71	Q8fu71 corynebacte
852	19	95.0	298	16	Q891M9	Q891m9 clostridium	925	19	95.0	311	16	Q88WP5	Q88wp5 lactobacill
853	19	95.0	299	2	Q9ADV6	Q9adv6 ehrlichia c	926	19	95.0	311	17	Q9HNX6	Q9hnx6 halobacteri
854	19	95.0	299	2	Q33HS7	Q33hs7 bradyrhizob	927	19	95.0	311	17	Q97ZS5	Q97zs5 sulfolobus
855	19	95.0	299	2	Q9GFG6	Q9gfg6 rhodococcus	928	19	95.0	312	2	Q8GFG9	Q8gfg9 rhodococcus
856	19	95.0	299	4	Q9NXA8	Q9nxa8 homo sapien	929	19	95.0	312	3	Q12125	Q12125 saccharomyc
857	19	95.0	299	4	Q9NXA8	Q9nx14 homo sapien	930	19	95.0	312	3	Q12125	Q12125 saccharomyc
858	19	95.0	299	4	Q9NXA8	Q9nx14 homo sapien	931	19	95.0	312	3	Q12125	Q12125 saccharomyc
859	19	95.0	299	4	Q9NXA8	Q9nx14 homo sapien	932	19	95.0	312	3	Q12125	Q12125 saccharomyc
860	19	95.0	299	4	Q9NXA8	Q9nx14 homo sapien	933	19	95.0	312	3	Q12125	Q12125 saccharomyc
861	19	95.0	299	11	Q9JHQ5	Q9jhg5 mus musculu	934	19	95.0	313	11	Q8VGE0	Q8vge0 mus musculu
862	19	95.0	299	11	Q8CPG8	Q8cdg8 mus musculu	935	19	95.0	313	11	Q8VGE0	Q8vge0 mus musculu
863	19	95.0	299	11	Q8BXX8	Q8brx8 mus musculu	936	19	95.0	313	16	Q92X69	Q92x69 rhizobium m
864	19	95.0	299	16	Q9KXN8	Q9kxw8 vibrio chol	937	19	95.0	313	16	Q92AX6	Q92ax6 listeria in
865	19	95.0	299	16	Q8EIG6	Q8eig6 vibrio chol	938	19	95.0	313	16	Q9Y6K0	Q9y6k0 listeria mo
866	19	95.0	299	16	Q8DSV7	Q8dsv7 shewanella	939	19	95.0	313	16	Q7U8G9	Q7u8g9 synechococc
867	19	95.0	299	16	Q87G20	Q87g20 vibrio para	940	19	95.0	314	5	Q8WQX2	Q8wqx2 leishmania
868	19	95.0	299	17	Q8ZWF0	Q8zwf0 pyrobaculum	941	19	95.0	314	11	Q8VFD7	Q8vfd7 mus musculu
869	19	95.0	299	17	Q8ZTV2	Q8ztv2 pyrobaculum	942	19	95.0	314	11	Q8VFD7	Q8vfd7 mus musculu
870	19	95.0	300	2	Q8GFG8	Q8gfg8 rhodococcus	943	19	95.0	314	11	Q7TRY1	Q7try1 mus musculu
871	19	95.0	300	16	Q8E3T8	Q8e3t8 streptococc	944	19	95.0	314	11	Q7TRY1	Q7try1 mus musculu
872	19	95.0	300	16	Q8D372	Q8d372 streptococc	945	19	95.0	314	11	Q7TRY1	Q7try1 mus musculu
873	19	95.0	300	16	Q7W5F1	Q7w5f1 bordetella	946	19	95.0	314	11	Q7TRY1	Q7try1 mus musculu
874	19	95.0	301	5	Q8IHK7	Q8ihk7 babesia div	947	19	95.0	314	11	Q7TRY1	Q7try1 mus musculu
875	19	95.0	301	10	Q93X14	Q93x14 oryza sativ	948	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
876	19	95.0	301	16	Q7VW47	Q7vw47 bordetella	949	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
877	19	95.0	301	17	Q8ZRP5	Q8zrp5 pyrobaculum	950	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
878	19	95.0	302	8	Q9SGH8	Q9sgn8 arabidopsis	951	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
879	19	95.0	302	10	Q9SGS4	Q9sgs4 arabidopsis	952	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
880	19	95.0	302	10	Q8LBJ9	Q8ldj9 arabidopsis	953	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
881	19	95.0	302	16	Q8KAL5	Q8kal5 chlorobium	954	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
882	19	95.0	302	16	Q8KAL5	Q8kal5 chlorobium	955	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
883	19	95.0	303	5	Q8M275	Q8m275 drosophila	956	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
884	19	95.0	303	5	Q8IPV4	Q8ipv4 drosophila	957	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
885	19	95.0	303	16	Q8NTU5	Q8ntu5 corynebacte	958	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
886	19	95.0	303	16	Q7WHF7	Q7whf7 bordetella	959	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
887	19	95.0	303	16	Q7W9B0	Q7w9b0 bordetella	960	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
888	19	95.0	304	2	Q33AC8	Q33ac8 pseudomonas	961	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
889	19	95.0	304	10	Q8LH88	Q8lbh8 arabidopsis	962	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
890	19	95.0	304	11	Q8BN03	Q8bn03 mus musculu	963	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
891	19	95.0	304	11	Q8P9J1	Q8p9j1 campylobact	964	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp
892	19	95.0	305	2	Q84EL1	Q84el1 stigmatalia	965	19	95.0	314	16	Q8YKH7	Q8ykh7 anabaena sp

```

966 19 95.0 317 9 Q854X2 mycobacteri
967 19 95.0 317 10 Q8LAG7 arabidopsis
968 19 95.0 317 11 Q97827 rattus norv
969 19 95.0 317 11 Q922J2 cricetus
970 19 95.0 317 16 Q812Y4 bacillus ce
971 19 95.0 317 17 Q30168 archaeglob
972 19 95.0 318 2 Q9X5F7 zymomonas m
973 19 95.0 318 2 Q83UD0 q83ud0 neisseria g
974 19 95.0 318 2 Q7WZ93 q7wz93 nonomurae
975 19 95.0 318 3 Q9C135 q9c135 mortierella
976 19 95.0 318 10 Q8LCB9 q8lcb9 arabidopsis
977 19 95.0 318 10 Q7XE27 q7xe27 oryza sativ
978 19 95.0 318 11 Q90Z21 q9gz21 mus musculu
979 19 95.0 318 11 Q8VGA9 q8vga9 mus musculu
980 19 95.0 318 16 Q8FEC0 q8fec0 escherichia
981 19 95.0 318 16 Q83LQ1 q83lq1 snigella fl
982 19 95.0 318 17 Q8TR71 q8tr71 methanosarc
983 19 95.0 318 17 Q8TQ14 q8tq14 methanosarc
984 19 95.0 319 3 Q9C0Z8 q9c0z8 schizosacch
985 19 95.0 319 5 Q95TC9 q95tc9 drosophila
986 19 95.0 319 10 Q9SW32 q9sw32 arabidopsis
987 19 95.0 319 16 Q8FTS9 q8fts9 corynebacte
988 19 95.0 319 16 Q83E34 q83e34 coxiella bu
989 19 95.0 319 17 Q9HK71 q9hk71 thermoplasma
990 19 95.0 319 17 Q8TR50 q8tr50 methanosarc
991 19 95.0 320 10 Q9LD47 q9ld47 arabidopsis
992 19 95.0 320 10 Q8LFT4 q8lft4 arabidopsis
993 19 95.0 320 10 Q8LKN3 q8lkn3 arabidopsis
994 19 95.0 320 10 Q9LUF3 q9luf3 arabidopsis
995 19 95.0 320 10 Q9C972 q9c972 arabidopsis
996 19 95.0 320 10 Q84WV9 q84wv9 arabidopsis
997 19 95.0 320 10 Q9CAH5 q9cah5 arabidopsis
998 19 95.0 321 2 Q7X1G1 q7x1g1 leptospiril
999 19 95.0 321 4 Q8NGA0 q8nga0 homo sapien
1000 19 95.0 322 2 Q7WX56 q7wx56 alcaligenes

```

## ALIGNMENTS

```

RESULT 1
>Q9TS18 PRELIMINARY; PRT; 20 AA.
ID Q9TS18;
AC Q9TS18;
>T 01-MAY-2000 (TREMBlrel. 13, Created)
>T 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
>T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
>E Cytotoxin-binding protein (Fragment).
>S Cryptococcus cuniculus (Rabbit).
>C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
>C Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
>X NCBI_TaxID=9986;
>R [1]
>R SEQUENCE FROM N.A.
>R MEDLINE=94039134; PubMed=7693466;
>R Lutz F., Mohr M., Grimmig M., Leidolf R., Linder D.;
>R "Pseudomonas aeruginosa cytotoxin-binding protein in rabbit
>R erythrocyte membranes. An oligomer of 28 kDa with similarity to
>R transmembrane channel proteins."
>R Eur. J. Biochem. 217:1123-1128(1993).
>R PIR: S39049; S39049.
>R HSSP: P47865; 134N.
>R GO: GO:0016020; C:membrane; IEA.
>R GO: GO:0005215; P:transporter activity; IEA.
>R GO: GO:0006810; P:transport; IEA.
>R InterPro: IPR000425; MIP.
>R Pfam: PF00230; MIP; 1.
>R SEQUENCE 20 AA; 2414 MW; 40B82D7DB5283D2D CRC64;

```

```

Query Match 95.0%; Score 19; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 VAEF 5
DB 14 VAEF 17

RESULT 2
Q9R5C0 PRELIMINARY; PRT; 22 AA.
ID Q9R5C0;
AC Q9R5C0;
>T 01-MAY-2000 (TREMBlrel. 13, Created)
>T 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
>T 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
>E Cytochrome P-460 (Fragment).
>S Nitrosomonas europaea.
>C Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
>C Nitrosomonadaceae; Nitrosomonas.
>X NCBI_TaxID=915;
>R [1]
>R SEQUENCE.
>R MEDLINE=93224468; PubMed=8385668;
>R McTavish H., LaQuier P., Arciero D., Logan M., Mundfrom G.,
>R Fuchs J.A., Hooper A.B.;
>R "Multiple copies of genes coding for electron transport proteins in
>R the bacterium Nitrosomonas europaea."
>R J. Bacteriol. 175:2445-2447(1993).
>R SEQUENCE 22 AA; 2549 MW; C22664F5EE1CE75F CRC64;

```

```

Query Match 95.0%; Score 19; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 VAEF 5
DB 3 VAEF 6

```

```

RESULT 3
P82214 PRELIMINARY; PRT; 30 AA.
ID P82214;
AC P82214;
>T 01-OCT-2001 (TREMBlrel. 18, Created)
>T 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
>T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
>E Myosin regulatory light chain 2 (MLC-2) (Fragment).
>S Bombyx mori (Silk moth).
>C Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
>C Neoptera; Endopterygota; Lepidoptera; Glossata; Bombycoidea;
>C Bombycidae; Bombyx.
>X NCBI_TaxID=7091;
>R [1]
>R SEQUENCE.
>R STRAIN=XINZHANG X KEMING; TISSUE=Body wall, and Fat body;
>R MEDLINE=21177481; PubMed=11280994;
>R Zhong B.X.;
>R "Protein database for several tissues derived from five instar of
>R silkworm."
>R I Chuan Hsueh Pao 28:217-224(2001).
>C -1- SUBUNIT: MYOSIN IS AN HEXAMER OF 2 HEAVY CHAINS & 4 LIGHT CHAINS
>C (BY SIMILARITY).
>C -1- MISCELLANEOUS: THIS CHAIN BINDS CALCIUM (BY SIMILARITY).
>C -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
>C GO: GO:0016459; F:calcium ion binding; IEA.
>C GO: GO:0005509; F:calcium development; IEA.
>C GO: GO:0007517; P:muscle development; IEA.
>C InterPro: IPR002048; EF-hand.
>C PROSITE: PS00018; EF HAND; PARTIAL.
>C MYOSIN; Calcium-binding; Muscle protein.
>C CA BIND 25 >30 BY SIMILARITY.
>C FT NON TER 30
>C SEQUENCE 30 AA; 3369 MW; 585287CE9EF538F3 CRC64;

```

```

Query Match 95.0%; Score 19; DB 5; Length 30;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;

```

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
13 VAEF 16

SULT 4  
5518

O25518 PRELIMINARY; PRT; 33 AA.  
01-JAN-1998 (TRENBLrel. 05, Created)  
01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
Hypothetical protein HP0847.  
Helicobacter pylori (Campylobacter pylori).  
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
Helicobacteraceae; Helicobacter.  
NCBI\_TaxID=210;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=26695 / ATCC 700392;  
MEDLINE=97394467; PubMed=9252185;  
Tomb J.-P., White O., Kierlavage A.R., Clayton R.A., Sutton G.G.,  
Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,  
Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,  
Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
Venter J.C.;  
"The complete genome sequence of the gastric pathogen Helicobacter  
pylori."  
Nature 388:539-547(1997).  
EMBL: AE000596; AAD07920.1; --  
PIR: G64625; G64625.  
TIGR: HP0847; --  
Hypothetical protein; Complete proteome.  
SEQUENCE 33 AA; 3704 MW; FA3F52631C0DB943 CRC64;  
Query Match 95.0%; Score 19; DB 16; Length 33;  
Best Local Similarity 100.0%; Pred. No. 5.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
15 VAEF 18

SULT 5  
ZZ19

Q8Z219 PRELIMINARY; PRT; 35 AA.  
01-MAR-2002 (TRENBLrel. 20, Created)  
01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
Hypothetical protein PAE0485.  
PAE0485.  
Pyrobaculum aerophilum.  
Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
Thermoproteaceae; Pyrobaculum.  
NCBI\_TaxID=13773;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=IM2 / ATCC 51768 / DSM 7523;  
MEDLINE=21664397; PubMed=11792869;  
Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
Miller J.H.;  
"Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
aerophilum."  
Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).

DR EMBL; AE009770; AAL62822.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 35 AA; 3934 MW; A324B4F94F6A14BB CRC64;  
Query Match 95.0%; Score 19; DB 17; Length 35;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
18 VAEF 21

Db

RESULT 6  
Q9TOR7 PRELIMINARY; PRT; 36 AA.  
ID Q9TOR7  
AC Q9TOR7;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Transferrin (Fragment).  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
OX NCBI\_TaxID=9796;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Giffard J.M., Brandon R.B., Bell T.K.;  
RT "Further identification of single nucleotide polymorphisms in the  
equine transferrin gene."  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF185729; AAF05508.1; --  
DR EMBL; AF185722; AAF05501.1; --  
DR EMBL; AF185723; AAF05502.1; --  
DR EMBL; AF185724; AAF05503.1; --  
DR EMBL; AF185725; AAF05504.1; --  
DR EMBL; AF185726; AAF05505.1; --  
DR EMBL; AF185727; AAF05506.1; --  
DR EMBL; AF185728; AAF05507.1; --  
DR HSSP; P18134; ITFD.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0008199; F:ferric iron binding; IEA.  
DR GO; GO:0006879; P:iron ion homeostasis; IEA.  
DR GO; GO:0006826; P:iron ion transport; IEA.  
DR InterPro; IPR001156; Transferrin.  
DR Pfam; PF00405; transferrin; 1.  
FT NON\_TER 1  
FT NON\_TER 36  
SQ SEQUENCE 36 AA; 3802 MW; E2DDD122186D5FA8 CRC64;  
Query Match 95.0%; Score 19; DB 6; Length 36;  
Best Local Similarity 100.0%; Pred. No. 5.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
28 VAEF 31

Db

RESULT 7  
Q48420 PRELIMINARY; PRT; 44 AA.  
ID Q48420  
AC Q48420;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
OS Klebsiella pneumoniae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Klebsiella.  
OX NCBI\_TaxID=573;  
RN [1]  
RP SEQUENCE FROM N.A.

```

C STRAIN=ATCC 25955;
A Willard B.L.;
T "Investigation of the Klebsiella pneumoniae 1,3-propanediol pathway:
T Characterization and expression of glycerol dehydratase and 1,3-
L propanediol oxidoreductase."
L Thesis (1994), Chemical Engineering, University of Wisconsin-Madison.
P [2]
P SEQUENCE FROM N.A.
C STRAIN=ATCC 25955;
A Skraly F.A., Willard B.L., Cameron D.C.;
T "The dha regulon of Klebsiella pneumoniae."
L Submitted (JUL-1995) to the EMBL/GenBank/DDJB databases.
C -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
R -1- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).
C EMBL: U30903; AAA74254.1; -.
R HSP; P11244; 1FX8.
R GO: GO:0016021; C:integral to membrane; IEA.
R GO: GO:0005215; F:transporter activity; IEA.
R GO: GO:0006910; P:transport; IEA.
R InterPro: IPR000425; MIP.
R Pfam: PF00230; MIP; 1.
R ProDom: PD000295; MIP family; 1.
W Hypothetical protein; Transmembrane; Transport.
T NON_TER 44 44
C SEQUENCE 44 AA; 4537 MW; 39951658FF89E734 CRC64;

Query Match 95.0%; Score 19; DB 2; Length 44;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
b 12 VAEF 15

RESULT 8
830H9 PRELIMINARY; PRT; 45 AA.
C Q830H9
T 01-JUN-2003 (TrEMBLrel. 24, Created)
T 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
E Hypothetical protein.
N E22805.
S Enterococcus faecalis (Streptococcus faecalis).
C Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
X NCBI_TaxID=1351;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=V583 / ATCC 700802;
MEDLINE=22550857; PubMed=12663927;
A Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
A Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
A Tettelin H., Dodson R.J., Unayam L., Brinkac L., Beanan M.,
A Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
A Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
A Utterback T., Ragune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
T "Role of mobile DNA in the evolution of vancomycin-resistant
T Enterococcus faecalis."
L Science 299:2071-2074 (2003).
R EMBL: A5016955; AA082500.1; -.
R TIGR: EF2805; -.
W Hypothetical protein; Complete proteome.
C SEQUENCE 45 AA; 5278 MW; 89CB55F5CB4C014A CRC64;

Query Match 95.0%; Score 19; DB 16; Length 45;
Best Local Similarity 100.0%; Pred. No. 6.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
b 21 VAEF 24

us-09-594-978a-3.rspt

RESULT 9
025478 PRELIMINARY; PRT; 48 AA.
ID O25478
AC O25478;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein HP0789.
GN HP0789.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547 (1997).
DR EMBL: AE000591; AAD07847.1; -.
DR FIR: E64618; E64618.
DR TIGR: HP0789; -.
KW Hypothetical protein; Complete proteome.
C SEQUENCE 48 AA; 5492 MW; ED5911D9F57BFEB CRC64;

Query Match 95.0%; Score 19; DB 16; Length 48;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
DB 15 VAEF 18

RESULT 10
Q7UXL5 PRELIMINARY; PRT; 51 AA.
ID Q7UXL5
AC Q7UXL5;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN R81255.
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1."
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL: BX294135; CAD71991.1; -.
KW Hypothetical protein; Complete proteome.
C SEQUENCE 51 AA; 5951 MW; 60F62389C0FDA6A7 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 51;

```

```

Best Local Similarity 100.0%; Pred. No. 7.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5
|||||
31 VAEF 34

SULT 11
4941
Q44941 PRELIMINARY; PRT; 52 AA.
Q44941;
01-NOV-1996 (TrEMBLrel. 01, Created)
01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
Short ORF.
LON
Bacillus brevis (Brevibacillus brevis).
Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
NCBI_TaxID=1393;
[1]
SEQUENCE FROM N.A.
STRAIN=HPD31;
MEDLINE=22388234; PubMed=12471157;
Ito K., Ueda S., Yamagata H.;
"Cloning, characterization, and inactivation of the Bacillus brevis
lon gene.";
J. Bacteriol. 174:2281-2287(1982).
EMBL; D00863; BAA00736.1; -.
PIR; A42375; A42375.
SEQUENCE 52 AA; 5558 MW; 5D689099F84F9AAF CRC64;

Query Match 95.0%; Score 19; DB 2; Length 52;
Best Local Similarity 100.0%; Pred. No. 7.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5
|||||
47 VAEF 50

SULT 12
X3Y8
Q8X3Y8 PRELIMINARY; PRT; 54 AA.
Q8X3Y8;
01-MAR-2002 (TrEMBLrel. 20, Created)
01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
Hypothetical protein z2382.
Z2382.
Escherichia coli O157:H7.
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
NCBI_TaxID=83334;
[1]
SEQUENCE FROM N.A.
STRAIN=O157:H7 / EDL933 / ATCC 700927;
MEDLINE=21074935; PubMed=11206551;
Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
Weich R.A., Blattner F.R.;
"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
Nature 409:529-533(2001).
EMBL; AE005369; AAC56425.1; -.
PIR; E85745; E85745.
Hypothetical protein; Complete proteome.
SEQUENCE 54 AA; 6507 MW; 8BEFACD8B56140CE CRC64;

Query Match 95.0%; Score 19; DB 16; Length 54;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 VAEF 5
|||||
DB 32 VAEF 35

RESULT 13
Q8FIF3 PRELIMINARY; PRT; 54 AA.
AC Q8FIF3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN C1469.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=417992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RL EMBL; AE016759; AAN79938.1; -.
DR EMBL; AE016759; AAN79938.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 54 AA; 5322 MW; 6B3CD0F21ED4376 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 54;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
|||||
DB 4 VAEF 7

RESULT 14
Q8VSE8 PRELIMINARY; PRT; 61 AA.
AC Q8VSE8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN C90186.
OS Shigella flexneri 2a.
OC Plasmid pCP301.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=42897;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301;
RA Jin Q., Zhang J.Y., Liu H., Yang J., Yang F., Zhang X.B., Wang J.H.,
RA Yang G.W., Wu H.T., Dong J., Sun L.L., Xue Y., Zhao A.L., Gao Y.S.,
RA Zhu J.P., Kan B., Chen S.X., Yao Z.J., He B.K., Chen R.S., Ma D.L.,
RA Yuan Z.H., Xu J.G., Wang Y., Shen Y., Lu W.C., Qiang B.Q., Wen Y.M.,
RA Hou Y.D.;
RT "Complete DNA sequence and analysis of the large virulence plasmid
pCP301 of Shigella flexneri";
Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF386526; AAL72547.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 61 AA; 6858 MW; F1CC17B10B28CBFC CRC64;

```

Query Match 95.0%; Score 19; DB 2; Length 61;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5  
 |||||  
 b 54 VAEF 57  
 |||||

RESULT 15  
 9P166 PRELIMINARY; PRT; 61 AA.  
 C Q3P166;  
 T 01-OCT-2000 (TrEMBLrel. 15, Created)  
 T 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 E PRO2435.  
 S Homo sapiens (Human).  
 C Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 X NCBI\_TaxID=9606;  
 N [1]  
 P SEQUENCE FROM N.A.  
 C TISSUE=Liver;  
 A Zhang C., Yu Y., Zhang S., Wei H., Zhang Y., Zhou G., Bi J., Liu M.,  
 A He F.;  
 I "Functional prediction of the coding sequences of 79 new genes deduced  
 I by analysis of cDNA clones from human fetal liver.";  
 L Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 R EMBL; AF119881; AAF69635.1; -  
 C SEQUENCE 61 AA; 7467 MW; 6433950D4657C1D4 CRC64;

Query Match 95.0%; Score 19; DB 4; Length 61;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5  
 |||||  
 b 30 VAEF 33  
 |||||

RESULT 16  
 99G29 PRELIMINARY; PRT; 68 AA.  
 C Q99G29;  
 I 01-JUN-2001 (TrEMBLrel. 17, Created)  
 I 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 I 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 E ORF49 (Hypothetical protein) (Unkown).  
 S Helicoverpa zea single nucleocapsid nucleopolyhedrovirus,  
 S Helicoverpa armigera nucleopolyhedrovirus G4, and  
 S Helicoverpa armigera nuclear polyhedrosis virus.  
 C Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 C Nucleopolyhedrovirus  
 X NCBI\_TaxID=10468, 148363, 51313;  
 N [1]  
 P SEQUENCE FROM N.A.  
 C SPECIES=Helicoverpa zea single nucleocapsid nucleopolyhedrovirus;  
 A Chen X., Zhang W.-J., Wong J., Chun G., Lu A., McCutchen B.F.,  
 A Presnail J.K., Herrmann R., Dolan M., Tingey S., Hu Z.-H., Vlak J.M.;  
 I "Genome sequence analysis of Helicoverpa zea single nucleocapsid  
 I nucleopolyhedrovirus.";  
 L Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 N [2]  
 P SEQUENCE FROM N.A.  
 C SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;  
 A Deng F., Chen X., Vlak J.M., Arif B.M., Hu Z.;  
 I "Sequence analysis of the gp37 gene of Heliothis armigera single-  
 I nucleocapsid nucleopolyhedrovirus.";  
 L Zhongguo Bingduxue 15:35-42(2000).  
 N [3]  
 P SEQUENCE FROM N.A.  
 C SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;

RA Wang H., Hu Z., Sun X., Vlak J.M., Chen X.;  
 RT "Sequence analysis of the iap3 gene of Heliothis armigera single-  
 RT nucleocapsid nucleopolyhedrovirus.";  
 RL Zhongguo Bingduxue 15:43-49(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;  
 RX MEDLINE=21078302; PubMed=11210934;  
 RA Wang H., Chen X., Wang H., Arif B.M., Vlak J.M., Hu Z.;  
 RT "Nucleotide sequence and transcriptional analysis of a putative basic  
 RT DNA-binding protein of Helicoverpa armigera polyhedrovirus.";  
 RL Virus Genes 22:113-120(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;  
 RX MEDLINE=21064569; PubMed=11125177;  
 RA Chen X., IJkel W.F., Tarchini R., Sun X., Sandbrink H., Wang H.,  
 RA Peters S., Zuidema D., Lankhorst R.K., Vlak J.M., Hu Z.;  
 RT "The sequence of the Helicoverpa armigera single-nucleocapsid  
 RT nucleopolyhedrovirus genome.";  
 RL J. Gen. Virol. 82:241-257(2001).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nucleopolyhedrovirus G4;  
 RX Chen X., IJkel W.F., Tarchini R., Sun X., Sandbrink H., Wang H.,  
 RA Peters S., Zuidema D., Lankhorst R.K., Vlak J.M., Hu Z.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nuclear polyhedrosis virus; STRAIN=C1;  
 RX PubMed=12050807;  
 RA Zhang C.X., Wu J.C.;  
 RT "Genome structure and the p10 gene of the Helicoverpa armigera  
 RT nucleopolyhedrovirus.";  
 RL Sheng Wu Hua Hsueh Yu Sheng Wu Li Hsueh Pao 33:179-184(2001).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nuclear polyhedrosis virus; STRAIN=C1;  
 RA Zhang C.X., Jin W.R.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Helicoverpa armigera nuclear polyhedrosis virus;  
 RA Fang M., Hu Z., Chen X., Vlak J.M.;  
 RT "Genetic organization of the HindIII-L region of Helicoverpa armigera  
 RT single-nucleocapsid nucleopolyhedrovirus.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF34030; AAL56194.1; -  
 DR EMBL; AF271059; AAG53791.1; -  
 DR EMBL; AF303045; AAK96298.1; -  
 DR EMBL; AF266694; AAK64316.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 68 AA; 7962 MW; 61B7718BFB195FF CRC64;

Query Match 95.0%; Score 19; DB 12; Length 68;  
 Best Local Similarity 100.0%; Pred. No. 1e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 |||||  
 DB 29 VAEF 32  
 |||||

RESULT 17  
 Q82X12 PRELIMINARY; PRT; 68 AA.  
 AC Q82X12;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Helix-turn-helix protein, Copg family.  
 GN NE0289.  
 OS Nitrosomonas europaea.

Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;  
Nitrosomonadaceae; Nitrosomonas.

NCBI\_TaxID=915;  
[1]

SEQUENCE FROM N.A.

STRAIN=ATCC 19718 / IF0 14298;

MEDLINE=22586410; PubMed=12700255;

Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,  
Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,  
Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;

"Complete genome sequence of the ammonia-oxidizing bacterium and  
obligate chemolithoautotroph Nitrosomonas europaea";

J. Bacteriol. 185:2759-2773(2003).

EMBL; BX321857; CAB64200.1; -.

Complete proteome.

SEQUENCE 68 AA; 7800 MW; 8A838B8537524DD6 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 68;

Best Local Similarity 100.0%; Pred. No. 1e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5

|||||

31 VAEF 34

SULT 18

X7L1

Q7X7L1 PRELIMINARY; PRT; 69 AA.

Q7X7L1;

01-OCT-2003 (TrEMBLrel. 25, Created)

01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

OSUNBA0044M19.22 protein [OSUNBA0053B21.1 protein].

OSUNBA0044M19.22 OR OSUNBA0053B21.1.

Oryza sativa (Rice).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzae; Oryza.

NCBI\_TaxID=4530;

[1]

SEQUENCE FROM N.A.

Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,

Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,

Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,

Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,

Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,

Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,

Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,

Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;

Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

EMBL; AL731601; CAB05035.1; -.

EMBL; AL731599; CAB05527.1; -.

SEQUENCE 69 AA; 7267 MW; 7783933BC87370BE CRC64;

Query Match 95.0%; Score 19; DB 10; Length 69;

Best Local Similarity 100.0%; Pred. No. 1e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5

|||||

31 VAEF 34

SULT 19

NPA7

Q8NPA7 PRELIMINARY; PRT; 69 AA.

Q8NPA7;

01-OCT-2002 (TrEMBLrel. 22, Created)

01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

Hypothetical protein Cg11907.

GN CG11907.  
OS Corynebacterium glutamicum (Brevibacterium flavum).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
CX NCBI\_TaxID=1718;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
RA Nakagawa S.;

"Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP005280; BAB99300.1; -.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 69 AA; 7815 MW; C0E2A072C2295DD2 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 69;

Best Local Similarity 100.0%; Pred. No. 1e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5

|||||

Db 23 VAEF 26

RESULT 20

Q8AUQ2 PRELIMINARY; PRT; 71 AA.

AC Q8AUQ2;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 23, Last annotation update)

DE Transferrin (Fragment).

OS Salmo trutta (Brown trout).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;

OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.

OC NCBI\_TaxID=8032;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Str-1, Str-2, Str-3, Str-4, Str-9, Str-10, and Str-11;

RX MEDLINE=2213592; PubMed=12140239;

RA Antunes A., Templeton A.R., Guyonard R., Alexandrino P.;

RT "The role of nuclear genes in intraspecific evolutionary inference:

genealogy of the transferrin gene in the brown trout.";

RL Mol. Biol. Evol. 19:1272-1287(2002).

DR EMBL; AF488850; AAN17027.1; -.

DR EMBL; AF488849; AAN17027.1; JOINED.

DR EMBL; AF488858; AAN17032.1; -.

DR EMBL; AF488857; AAN17032.1; JOINED.

DR EMBL; AF488866; AAN17037.1; -.

DR EMBL; AF488865; AAN17037.1; JOINED.

DR EMBL; AF488874; AAN17042.1; -.

DR EMBL; AF488873; AAN17042.1; JOINED.

DR EMBL; AF488914; AAN17068.1; -.

DR EMBL; AF488913; AAN17068.1; JOINED.

DR EMBL; AF488922; AAN17073.1; -.

DR EMBL; AF488921; AAN17073.1; JOINED.

DR EMBL; AF488930; AAN17078.1; -.

DR EMBL; AF488929; AAN17078.1; JOINED.

DR GO; GO:0005576; C:extracellular; IEA.

DR GO; GO:0008199; F:ferrous iron binding; IEA.

DR GO; GO:0008799; P:iron ion homeostasis; IEA.

DR GO; GO:0008826; P:iron ion transport; IEA.

DR InterPro; IPR001156; Transferrin.

DR Pfam; PF00405; Transferrin; 1.

DR PRINTS; PR00422; TRANSFERRIN.

FT NON TER 1 71

FT NON TER 71 71

SQ SEQUENCE 71 AA; 7546 MW; 892E38F22D2AC4D CRC64;

Query Match 95.0%; Score 19; DB 13; Length 71;

Best Local Similarity 100.0%; Pred. No. 1.1e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Y      2 VAEF 5
b      33 VAEF 36

RESULT 21
8AUP0 PRELIMINARY; PRT; 71 AA.
D Q8AUP0
T 01-MAR-2003 (TREMELrel. 23, Created)
T 01-MAR-2003 (TREMELrel. 23, Last sequence update)
T 01-OCT-2003 (TREMELrel. 25, Last annotation update)
E Transferin (Fragment).
S Salmo salar (Atlantic salmon).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Actinopterygii; Neopterygii; Teleostei; Euteleostei;
C Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
X NCBI_TaxID=8030;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=Sea-1, and Ssa-2;
X MEDLINE=2213592; PubMed=12140239;
A Antunes A., Templeton A.R., Guyomard R., Alexandrino P.;
T "The role of nuclear genes in intraspecific evolutionary inference:
T genealogy of the transferrin gene in the brown trout.";
L Mol. Biol. Evol. 19:1272-1287(2002).
R EMBL; AF488834; AAN17017.1; -.
R EMBL; AF488833; AAN17017.1; JOINED.
R EMBL; AF488842; AAN17022.1; -.
R EMBL; AF488841; AAN17022.1; JOINED.
R GO: GO:0005576; C:extracellular; IEA.
R GO: GO:0008199; F:ferric iron binding; IEA.
R GO: GO:0008799; P:iron ion homeostasis; IEA.
R GO: GO:0006826; P:iron ion transport; IEA.
R InterPro; IPR001156; Transferrin.
R Pfam; PF00405; Transferrin; 1.
R PRINTS; PR00422; TRANSFERRIN.
R T NON_TER 1
R NON_TER 1
T SEQUENCE 71 AA; 7504 MW; 90886A1234CBAC4D CRC64;

Query Match 95.0%; Score 19; DB 13; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      2 VAEF 5
b      33 VAEF 36

RESULT 22
'97252 PRELIMINARY; PRT; 72 AA.
D P97252
C P97252
T 01-MAY-1997 (TREMELrel. 03, Created)
T 01-MAY-1997 (TREMELrel. 03, Last sequence update)
T 01-MAY-2000 (TREMELrel. 13, Last annotation update)
E Late control gene D protein (Fragment).
N D.
N Escherichia coli.
X Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
X Enterobacteriaceae; Escherichia.
X NCBI_TaxID=562;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=KJ2;
X MEDLINE=97251358; PubMed=9097040;
A Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
A Kasai H., Kimura S., Kitagawa M., Kitagawa M., Makino K., Maki T.,
A Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
A Nishimoto H., Nishio Y., Oshima T., Saito N., Sampa G., Seki Y.,
A Sivasaundaram S., Tagami H., Takeda J., Takemoto K., Wada C.,

Yamamoto Y., Horiuchi T.;
RA "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
DR EMBL; D90847; BAA15949.1; -.
DR EMBL; D90846; BAA15938.1; -.
FT NON_TER 1
SQ SEQUENCE 72 AA; 8099 MW; 42A63B25B00EADCB CRC64;

Query Match 95.0%; Score 19; DB 2; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      2 VAEF 5
b      3 VAEF 6

RESULT 23
Q97JU4 PRELIMINARY; PRT; 74 AA.
ID Q97JU4
AC Q97JU4;
DT 01-OCT-2001 (TREMELrel. 18, Created)
DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein CAC1179.
GN CAC1179.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=2135925; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatousov R.I., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AS007632; XAK79151.1; -.
DR PIR; D97045; D97045.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 74 AA; 8747 MW; 633633CB0A0C293A CRC64;

Query Match 95.0%; Score 19; DB 16; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      2 VAEF 5
b      5 VAEF 8

RESULT 24
Q82MX0 PRELIMINARY; PRT; 74 AA.
ID Q82MX0
AC Q82MX0;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein.
GN SAV1532.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;

```

Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,  
Kikuchi H., Shiba T., Sakaki Y., Hattori M.,  
"Genome sequence of an industrial microorganism Streptomyces  
avermitilis: deducing the ability of producing secondary  
metabolites";  
Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).

[2]  
SEQUENCE FROM N.A.

STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
MEDLINE=22608306; PubMed=12692562;  
Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
Sakaki Y., Hattori M., Omura S.;  
"Complete genome sequence and comparative analysis of the industrial  
microorganism Streptomyces avermitilis";  
Nat. Biotechnol. 21:526-531(2003).

EMBL; AP005027; BAC69243.1; -

Hypothetical protein; Complete proteome.

SEQUENCE 74 AA; 7960 MW; D8CE71C8B08AEB01 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 74;

Best Local Similarity 100.0%; Pred. NO. 1.1e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5

|||||

36 VAEF 39

SULT 25

62M8

Q862M8 PRELIMINARY; PRT; 77 AA.

01-JUN-2003 (TRENBLrel. 24, Created)

01-JUN-2003 (TRENBLrel. 24, Last sequence update)

01-OCT-2003 (TRENBLrel. 25, Last annotation update)

Similar to MAD2 protein (fragment).

Bos taurus (Bovine).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

NCBI\_TaxID=9913;

[1]

SEQUENCE FROM N.A.

MEDLINE=22544902; PubMed=12658628;

Ishiwata H., Katsuma S., Kizaki K., Patel O.V., Nakano H.,

Takahashi T., Inai K., Hirasawa A., Shiojima S., Ikawa H., Suzuki Y.,

Tsujimoto G., Izaike Y., Todoroki J., Hashizume K.;

"Characterization of gene expression profiles in early bovine

pregnancy using a custom cDNA microarray";

Mol. Reprod. Dev. 65:9-18(2003).

EMBL; AB098952; BAC56442.1; -

InterPro; IPR003511; DNABind\_HORMA.

PROSITE; PS50815; HORMA; 1.

NON\_TER 1

NON\_TER 77

SEQUENCE 77 AA; 8832 MW; E2E1D625C848E571 CRC64;

Query Match

Best Local Similarity 95.0%; Score 19; DB 6; Length 77;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5

|||||

6 VAEF 9

SULT 26

18VS7

Q88VS7 PRELIMINARY; PRT; 77 AA.

01-JUN-2003 (TRENBLrel. 24, Created)

01-JUN-2003 (TRENBLrel. 24, Last sequence update)

DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)

DE Hypothetical protein.

GN LP 1960.

OS Lactobacillus plantarum.

OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;

OC Lactobacillus.

OX NCBI\_TaxID=1590;

[1]

SEQUENCE FROM N.A.

STRAIN=NCIMB 8826 / WCFS1;

MEDLINE=22480296; PubMed=12566566;

Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,

Kuipers O.P., Leer R., Turchini R., Peters S.A., Sandbrink H.M.,

Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,

Hoffers S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ussing B.,

De Vos W.M., Siezen R.J.;

"Complete genome sequence of Lactobacillus plantarum WCFS1";

Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).

EMBL; AL935257; CAD64344.1; -

Hypothetical protein; Complete proteome.

SEQUENCE 77 AA; 8618 MW; C00B2BFF1D401F2F CRC64;

Query Match 95.0%; Score 19; DB 16; Length 77;

Best Local Similarity 100.0%; Pred. NO. 1.2e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5

|||||

68 VAEF 71

RESULT 27

Q98582 PRELIMINARY; PRT; 79 AA.

AC Q98582; 1997 (TRENBLrel. 02, Created)

DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)

DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)

DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)

DE A532L protein.

GN A532L

OS Paramesidium bursaria chlorella virus 1 (PBCV-1).

OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.

OX NCBI\_TaxID=10506;

[1]

SEQUENCE FROM N.A.

RX MEDLINE=36400190; PubMed=8806566;

RA Kurish G.F., Li Y., Lu Z., Furuta M., Rock D.L., Van Etten J.L.;

"Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map

positions 182 to 258.";

RL Virology 223:303-317(1996).

[2]

SEQUENCE FROM N.A.

RX MEDLINE=20013326; PubMed=10544099;

RA Kaiser A., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,

Lisea A.D., Nickerson K.W., Van Etten J.L.;

"Chlorella virus PBCV-1 encodes a functional homospesmidine

synthase.";

RL Virology 263:254-262(1999).

[3]

SEQUENCE FROM N.A.

RX MEDLINE=20478054; PubMed=11021991;

RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;

"Characterization of a beta-1,3-glucanase encoded by chlorella virus

PBCV-1.";

RL Virology 276:27-36(2000).

[4]

SEQUENCE FROM N.A.

RA Van Etten J.L.;

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

[5]

SEQUENCE FROM N.A.

RA Van Etten J.L.;

RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.

N [6]  
P SEQUENCE FROM N.A.  
A Van Etten J.L.;  
L Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
N  
P SEQUENCE FROM N.A.  
A Graves M.V., Van Etten J.L.;  
L Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
N  
P SEQUENCE FROM N.A.  
A Graves M.V., Van Etten J.L.;  
L Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
N  
P SEQUENCE FROM N.A.  
A Gurnon J.R., Graves M.V., Van Etten J.L.;  
L Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
N  
P EMBL; U42580; AAC96899.1; -  
R PIR; T18034; T18034.  
Q SEQUENCE 79 AA; 8698 MW; B191C627F5D5CSA7 CRC64;

Query Match 95.0%; Score 19; DB 12; Length 79;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5  
|||  
b 49 VAEF 52

RESULT 28  
9KI36 PRELIMINARY; PRT; 80 AA.  
D Q9KI36  
C Q9KI36;  
T 01-OCT-2000 (TrEMBLrel. 15, Created)  
T 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
E Ysg.  
N Ysg.  
S Agrobacterium tumefaciens.  
G Plasmid Ti.  
C Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
C Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
X NCBI\_TaxID=358;  
N [1]  
P SEQUENCE FROM N.A.  
A Winans S.C., Zhu J., Oger P.M., Schrammeijer B., Hooykaas P.J.,  
A Farrand S.K.;  
T "Octopine-type Ti plasmid sequence."  
L Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
R EMBL; AF242681; AAF77180.1; -  
R GO; GO:0046821; C:extrachromosomal DNA; IEA.  
W Plasmid.  
Q SEQUENCE 80 AA; 8723 MW; 4B8379C1D907CA8E CRC64;

Query Match 95.0%; Score 19; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5  
|||  
b 65 VAEF 68

RESULT 29  
J89WL6 PRELIMINARY; PRT; 81 AA.  
D Q89WL6  
C Q89WL6;  
T 01-JUN-2003 (TrEMBLrel. 24, Created)  
T 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
E Bsr0662 protein.  
N Bsr0662.  
S Bradyrhizobium japonicum.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bradyrhizobiaceae; Bradyrhizobium.  
RN NCBI\_TaxID=375;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=USDA 110;  
RX MEDLINE=22484998; PubMed=12597275;  
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
RA Sasamoto S., Watanabe A., Idesawa K., Iriiguchi M., Kawashima K.,  
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,  
RA Tabata S.;  
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110.";  
RL DNA Res. 9:189-197 (2002).  
DR EMBL; AP005937; BAC45927.1; -  
KW Complete proteome.  
SQ SEQUENCE 81 AA; 9235 MW; D512A9FFD0DA6C7 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
|||  
DB 43 VAEF 46

RESULT 30  
Q9KAV4 PRELIMINARY; PRT; 84 AA.  
ID Q9KAV4  
AC Q9KAV4;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein BH2182.  
GN BH2182.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
Nucleic Acids Res. 28:4317-4331 (2000).  
RL EMBL; AF001514; BAB05901.1; -  
DR PIR; F83922; F83922.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 84 AA; 9669 MW; 6CDE3768ED3F5D84 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 84;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
|||  
DB 38 VAEF 41

RESULT 31  
Q821D2 PRELIMINARY; PRT; 84 AA.  
ID Q821D2  
AC Q821D2;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN CC801010.  
OS Chlamydomonadales.



```

Y      2 VAEF 5
D      13 VAEF 16

RESULT 35
7Y3Y9 PRELIMINARY; PRT; 87 AA.
D Q7Y3Y9
I 01-OCT-2003 (TREMBlrel. 25, Created)
I 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
I 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
E Hypothetical protein.
S Bacteriophage PY54.
S Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
X NCBI_TaxID=172667;
X [1]
P SEQUENCE FROM N.A.
A Hertwig S., Klein I., Schmidt V., Beck S., Hammerl J.A., Appel B.;
I "Sequence analysis of the genome of the temperate Yersinia
I enterocolitica phage PY54.";
I Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
Y [2]
P SEQUENCE FROM N.A.
A Hertwig S.;
I Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
L EMBL; AJ564013; CAD91786.1; -.
N Exonuclease; Hypothetical protein.
Q SEQUENCE 87 AA; 8538 MW; D17D7A7E3075459B CRC64;

Query Match 95.0%; Score 19; DB 9; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Y      2 VAEF 5
b     31 VAEF 34

RESULT 36
98590 PRELIMINARY; PRT; 88 AA.
D Q98590
I 01-MAY-2000 (TREMBlrel. 13, Created)
I 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
I 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
E Cytochrome P460 (Fragment).
N CYP.
S Nitrosomonas europaea.
C Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
C Nitrosomonadaceae; Nitrosomonas.
X NCBI_TaxID=915;
X [1]
P SEQUENCE FROM N.A.
A Iizumi T., Nakamura K.;
I "Construction of tryptophan requiring-mutant of Nitrosomonas europaea
I by inactivation of trpC gene using homologous recombination.";
I Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
L EMBL; AB030031; BAA83388.1; -.
T NON TER 88
Q SEQUENCE 88 AA; 9765 MW; A59003348F4E280C CRC64;

Query Match 95.0%; Score 19; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Y      2 VAEF 5
b     29 VAEF 32

```

```

RESULT 37
Q41185 PRELIMINARY; PRT; 88 AA.
ID Q41185
AC Q41185;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Glyceraldehyde-3-phosphate-dehydrogenase subunit GapB (Fragment).
GN GAPB.
OS Arabidopsis thaliana (Mouse-ear cress).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93013005; PubMed=1398114;
RA Shih M.C., Heinrich P., Goodman H.M.;
RT "Cloning and chromosomal mapping of nuclear genes encoding chloroplast
RT and cytosolic glyceraldehyde-3-phosphate-dehydrogenase from
RT Arabidopsis thaliana.";
RL Gene 119:317-319(1992).
DR EMBL; S45911; AAB23533.1; -.
DR GO; GO:0009507; C:chloroplast; IEA.
KW Chloroplast.
FT NON TER 88
SQ SEQUENCE 88 AA; 9198 MW; FCD56A7E37B6999D CRC64;

Query Match 95.0%; Score 19; DB 10; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 VAEF 5
Db     38 VAEF 41

RESULT 38
Q9ZGY2 PRELIMINARY; PRT; 89 AA.
ID Q9ZGY2
AC Q9ZGY2;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN Y1103.
OS Yersinia pestis.
OG Plasmid pMT-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
CX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM10+.
RX MEDLINE=99043698; PubMed=9826348;
RA Lindler L.E., Plano G.V., Burland V., Mayhew G.F., Blattner F.R.;
RT "Complete DNA sequence and detailed analysis of the Yersinia pestis
RT KIM5 plasmid encoding murine toxin and capsular antigen.";
RL Infect. Immun. 66:5731-5742(1998).
DR EMBL; AF074611; AAC82761.1; -.
DR FR; T15018; T15018.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR Interpro; IPR003006; IG_MHC.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 89 AA; 10736 MW; 72B4221126023EE4 CRC64;

Query Match 95.0%; Score 19; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 VAEF 5

```

Query Match 95.0%; Score 19; DB 16; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0

QY 2 VAEF 5  
DB 68 VAEF 71

RESULT 41  
Q81ZX3 PRELIMINARY; PRT; 90 AA.  
ID Q81ZX3  
AC Q81ZX3  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE MAD2 mitotic arrest deficient-like 1 variant.  
GN MAD2L1  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yin F., Fan D.M.;  
RT "Identifying a new variant of MAD2L1";  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF394735; AAN74648.1;  
DR InterPro; IPR003511; DNABind\_HORMA.  
DR Pfam; PF02301; HORMA; 1.  
DR PROSITE; PS00815; HORMA; 1.  
SQ SEQUENCE 90 AA; 10335 MW; 8209F5A7AD8D09B CRC64;

Query Match 95.0%; Score 19; DB 4; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0

QY 2 VAEF 5  
DB 20 VAEF 23

RESULT 42  
Q98TB2 PRELIMINARY; PRT; 91 AA.  
ID Q98TB2  
AC Q98TB2  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Preproinsulin (Fragment).  
OS Ambloplites rupestris (Rock bass).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Percoidae;  
OC Centrarchidae; Ambloplites.  
OX NCBI\_TaxID=109273;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Al-Mahrouki A.A., Irwin D.M., Youson J.H.;  
RT "Molecular cloning of preproinsulin cDNA from the rock bass";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
DR EMBL; AF199584; AAK26708.1;  
DR HSP; P01308; ILPH.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0005179; F:hormone activity; IEA.  
DR GO; GO:0007582; P:physiological processes; IEA.  
DR InterPro; IPR004825; Ins/IGF/relax.  
DR Pfam; PF00049; Insulin; 1.  
DR PRINTS; PR00277; INSULINB.  
DR SMART; SM00078; ILGF; 1.

SULT 39  
CNX1 PRELIMINARY; PRT; 89 AA.  
Q9CNX1  
Q9CNX1  
01-JUN-2001 (TREMBlrel. 17, Created)  
01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
RPS15.  
RPS15 OR PM0301.  
Pasteurella multocida.  
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
Pasteurellaceae; Pasteurella.  
NCBI\_TaxID=747;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=PM70;  
MEDLINE=21145866; PubMed=11248100;  
May B.-J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
"Complete genomic sequence of Pasteurella multocida PM70";  
Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).  
EMBL; AE006066; AAK02385.1; --  
HSP; P80378; LAB3.  
GO; GO:0005622; C:intracellular; IEA.  
GO; GO:0005840; C:ribosome; IEA.  
GO; GO:0003735; F:structural constituent of ribosome; IEA.  
GO; GO:0006412; P:protein biosynthesis; IEA.  
InterPro; IPR000589; Ribosomal\_S15.  
InterPro; IPR005290; Ribosomal\_S15\_b.  
Pfam; PF00312; Ribosomal\_S15; 1.  
Prodom; PD157043; RPS15\_bact; 1.  
TIGRFAMs; TIGR00952; S15\_bact; 1.  
PROSITE; PS00362; RIBOSOMAL\_S15; 1.  
Complete proteome.  
SEQUENCE 89 AA; 10184 MW; F796E0AB5283ED0E CRC64;

Query Match 95.0%; Score 19; DB 16; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0

QY 2 VAEF 5  
DB 12 VAEF 15

RESULT 40  
Q9CFD1 PRELIMINARY; PRT; 89 AA.  
ID Q9CFD1  
AC Q9CFD1  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Hypothetical protein YPJ8.  
OS Yersinia enterocolitica (subsp. biovar 4/O:3).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1360;  
RN [1]  
RP SEQUENCE FROM N.A.  
STRAIN=IL1403;  
MEDLINE=21235196; PubMed=11337471;  
Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,  
Weissenbach J., Ehrlich S.D., Sorokin A.;  
"The complete genome sequence of the lactic acid bacterium Lactococcus  
lactis ssp. lactis IL1403";  
Genome Res. 11:731-753 (2001).  
EMBL; AE006385; AAK05648.1; --  
FIR; F86818; F86818.  
Hypothetical protein; Complete proteome.  
SEQUENCE 89 AA; 10403 MW; D90DF3CF71D12D06 CRC64;

DR PROSITE; PS00262; INSULIN; 1.  
 FT NON TER 1  
 FT NON TER 91  
 SQ SEQUENCE 91 AA; 10100 MW; E86C8B256DC6D939 CRC64;

Query Match 95.0%; Score 19; DB 13; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 55 VAEF 58

## RESULT 43

8KV12 ID Q8KY12 PRELIMINARY; PRT; 94 AA.

AC Q8KY12; 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Putative transcriptional regulator (Fragment).  
 OS Rhizobium etli.  
 NCBI\_TaxID=29449;

GN Plasmid p42b.  
 OS Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
 OX NCBI\_TaxID=29449;

RP SEQUENCE FROM N.A.  
 STRAIN=CE3;

LA Cevallos M.A., Izquierdo J., Porta H., Tun C., Davila G., Brom S.;  
 "Rhizobium etli CE3 contains at least three plasmids of the RepABC  
 family: A structural and an evolutionary analysis";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF313446; ANW88940.1; -  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.

FT NON TER 1  
 SQ SEQUENCE 94 AA; 10222 MW; C09D6EA3DFF1E2BA CRC64;

Query Match 95.0%; Score 19; DB 2; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 38 VAEF 41

## RESULT 44

046425 ID Q46425 PRELIMINARY; PRT; 94 AA.

AC Q46425; 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Aquaporin 1 (Fragment).

OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;

RP SEQUENCE FROM N.A.  
 STRAIN=New Zealand;

LA Carter E.P., Umenishi F., Matthay M.A., Verkman A.S.;  
 "Increased water permeability across the blood-gas barrier in rabbit  
 lungs in the first 24 hours after birth";  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -|- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).  
 DR EMBL; AF000311; AAB94408.1; -  
 DR HSP; P29972; 1FOY.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0005741; C:mitochondrial outer membrane; IEA.  
 DR GO; GO:0015288; F:porin activity; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR000425; MIP.  
 DR Pfam; PF00230; MIP; 1.  
 DR PRODOM; PD000295; MIP\_family; 1.  
 DR PROSITE; PS00221; MIP; 1.  
 KW Porin; Transmembrane; Transport.

FT NON TER 1  
 FT NON TER 94  
 SQ SEQUENCE 94 AA; 10057 MW; 963D5527631E8CDC CRC64;

Query Match 95.0%; Score 19; DB 6; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 9 VAEF 12

## RESULT 45

Q8U8Q1 ID Q8U8Q1 PRELIMINARY; PRT; 96 AA.

AC Q8U8Q1; 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Hypothetical protein Atu4039.

GN Atu4039.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
 OX NCBI\_TaxID=176299;

RP SEQUENCE FROM N.A.

RA MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura V.K., Zhou Y., Chen L., Wood G.B., Almeida N.F. Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,  
 RA Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutayin T., Levy R., Li M.-J., McClelland E., Palmieri A.,  
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 C58";  
 RL Science 294:2317-2323 (2001).  
 DR EMBL; AE009334; AAL44840.1; -  
 DR PIR; AB3053; AB3053.

KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 96 AA; 11193 MW; FBD635894B46A8AE CRC64;

Query Match 95.0%; Score 19; DB 16; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 DB 55 VAEF 58

## RESULT 46

Q89WX7 ID Q89WX7 PRELIMINARY; PRT; 96 AA.

AC Q89WX7; 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Bsl0551 protein.

BSL0551.  
Bradyrhizobium japonicum.  
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
Bradyrhizobiaceae; Bradyrhizobiaceae;  
NCBI\_TaxID=375;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=USDA 110;  
MEDLINE=22484998; PubMed=12597275;  
Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
Sasamoto S., Watanabe A., Idegawa K., Iriuchi M., Kawashima K.,  
Kohara M., Matsumoto M., Shimpō S., Tezuka H., Wada T., Yamada M.,  
Tabata S.;  
"Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110.";  
DNA Res. 9:189-197(2002).  
EMBL: AP005936; BAC45816.1; -;  
GO: GO:0016020; C:membrane; IEA.  
InterPro: IPR003425; Unk\_YGGT.  
Pfam: PF02325; YGGT; 1.  
Complete proteome.  
SEQUENCE 96 AA; 10976 MW; A9E11F9AA4BC7734 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 96;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
|||||  
44 VAEF 47

SULT 47  
UO09

Q8U0G9 PRELIMINARY; PRT; 96 AA.  
Q8U0G9;  
01-JUN-2002 (TrEMBLrel. 21, Created)  
01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
Hypothetical protein PF1620.  
PF1620.  
Pyrococcus furiosus.  
Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
Pyrococcus.  
NCBI\_TaxID=2261;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=Vc1 / DSM 3638 / ATCC 43597 / JCM 8422;  
Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
"The complete sequence of the Pyrococcus furiosus genome.";  
Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
EMBL: AB010282; AA181744.1; -;  
Hypothetical protein; Complete proteome.  
SEQUENCE 96 AA; 10804 MW; D6DAB09D096D577A CRC64;

Query Match 95.0%; Score 19; DB 17; Length 96;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
|||||  
24 VAEF 27

SULT 48  
6535

Q86535 PRELIMINARY; PRT; 97 AA.  
Q86535;  
01-NOV-1996 (TrEMBLrel. 01, Created)  
01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
2C/3A (Fragment).  
2C/3A.

OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
[1]  
RN SEQUENCE FROM N.A.  
RP  
RC STRAIN=LSH/8;  
RX MEDLINE=92348853; PubMed=1668326;  
RA Fineschi N., Cavalieri F., Garelick H., Prugnola A., Pellegrini V.,  
Zuckerman A.J.;  
RT "Characterization of a hepatitis A virus strain suitable for vaccine  
production.";  
J. Hepatol. 13:S146-S151(1991).  
RL EMBL: S44109; AAB22740.2; -;  
FT NON TER 1 1  
FT NON TER 97 97  
SQ SEQUENCE 97 AA; 10614 MW; D207F7CB93110DDF CRC64;

Query Match 95.0%; Score 19; DB 12; Length 97;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
|||||  
44 VAEF 47

Db

RESULT 49  
Q8Y476 PRELIMINARY; PRT; 97 AA.  
ID Q8Y476  
AC Q8Y476;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein lmo2579.  
GN LMO2579.  
OS Listeria monocytogenes.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OC NCBI\_TaxID=1639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EGD-e / Serovar 1/2a;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,  
Entian K.-D., Fsihi H., Garcia-del Portillo F., Garido P.,  
Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hain J., Jackson D.,  
Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,  
Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
Vaquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
"Comparative genomics of Listeria species.";  
Science 294:849-852(2001).  
RL EMBL: AL591983; CAD00657.1; -;  
DR PIR: AC1397; AC1397.  
DR Listalist; LMO02579; -;  
DR InterPro: IPR007138; ABM.  
DR Pfam: PF03932; ABM; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 97 AA; 10379 MW; 9E758586E94218E0 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 97;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
|||||  
15 VAEF 18

Db

```

RESULT 50
9H238 PRELIMINARY; PRT; 99 AA.
C Q9H238;
C Q9H238;
T 01-MAR-2001 (TrEMBLrel. 16, Created)
T 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
E Hypothetical protein PA3202.
N PA3202.
S Pseudomonas aeruginosa.
C Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
C Pseudomonadaceae; Pseudomonas.
K NCBI_TaxID=287;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=ATCC 15692 / PA01;
C MEDLINE=20437337; PubMed=10984043;
A Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
A Hickey M.J., Pham X.-Q.T., Hufnagle W.O., Kowalik D.J., Lagrou M.,
A Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
A Brody L.D., Coulter S.N., Folger K.R., Kas A., Laibig K., Lim R.M.,
A Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
A Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
I "Complete genome sequence of Pseudomonas aeruginosa PA01, an
I opportunistic pathogen.";
L Nature 406:959-964(2000).
R EMBL; AE004744; AG08590.1; -.
R PIR; E83244; E83244.
R InterPro; IPR005545; YCII.
R Pfam; PF03795; YCII; 1.
W Hypothetical protein; Complete proteome.
Q SEQUENCE 99 AA; 10609 MW; 3AD945F4D54A85C CRC64;

Query Match 95.0%; Score 19; DB 16; Length 99;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
D 62 VAEF 55

RESULT 51
885M3 PRELIMINARY; PRT; 99 AA.
C Q885M3;
C Q885M3;
T 01-JUN-2003 (TrEMBLrel. 24, Created)
T 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E Conserved hypothetical protein.
N PSPT01808.
S Pseudomonas syringae (pv. tomato).
C Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
C Pseudomonadaceae; Pseudomonas.
X NCBI_TaxID=323;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=DC3000;
C Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
C Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
C Dodson R., DeBey R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
C Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidson T.,
C White O., Fraser C., Collier A.;
I "Complete sequence of Pseudomonas syringae.";
L Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
R EMBL; AE018662; AAO5328.1; -.
R TIGR; PSPT01808; -.
R InterPro; IPR005545; YCII.
R Pfam; PF03795; YCII; 1.
W Hypothetical protein; Complete proteome.
Q SEQUENCE 99 AA; 10504 MW; 62836007E4849392 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 99;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
D 62 VAEF 55

RESULT 52
Q82UR2 PRELIMINARY; PRT; 100 AA.
ID Q82UR2;
AC Q82UR2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN NE1419.
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19718 / IFO 14298;
RX MEDLINE=22586410; PubMed=12700255;
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Savavedra-Soto L.A.,
RA Arciero D.M., Hommes N.G., Whittaker M.M., Alp D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
RT obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773(2003).
DR EMBL; BX321861; CAD85330.1; -.
DR InterPro; IPR005545; YCII.
DR Pfam; PF03795; YCII; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 100 AA; 10836 MW; 82DC1153B5ABFAE27 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
D 62 VAEF 65

RESULT 53
Q8VS54 PRELIMINARY; PRT; 101 AA.
ID Q8VS54;
AC Q8VS54;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Elongation factor TS (fragment).
GN TSF
OS Lactobacillus reuteri.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1598;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 55739;
RA Nam S.J., Kim J.K., Park J.Y., Ha Y.L., Kim J.H.;
RT "Cloning of UMP-kinase gene from Lactobacillus reuteri ATCC 557939.";
L Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
R EMBL; AF401482; AAL60142.1; -.
R DR GO; GO:0003746; P:translation elongation factor activity; IEA.
R DR GO; GO:0006414; P:translational elongation; IEA.
R InterPro; IPR001816; EF_TS.
R Pfam; PF00889; EF_TS; 1.
R NON_TER 1
SQ SEQUENCE 101 AA; 11650 MW; B5CF30D941DE2B9B CRC64;
```

Query Match 95.0%; Score 19; DB 2; Length 101;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
 ||||  
 62 VAEF 65

SULT 54  
 7YC2 PRELIMINARY; PRT; 101 AA.

Q97YC2  
 01-OCT-2001 (TReMBLrel. 18, Created)  
 01-OCT-2001 (TReMBLrel. 18, Last sequence update)  
 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 Hypothetical protein SSO1404.  
 SSO1404.  
 Sulfolobus solfataricus.  
 Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 Sulfolobus.  
 NCBI\_TaxID=2287;  
 [1]

SEQUENCE FROM N.A.  
 STRAIN=ATCC 35092 / DSM 1617 / P2;  
 MEDLINE=21332296; PubMed=11427726;  
 She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 Avez M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
 De Moors A., Erasus G., Fletcher C., Gordon P.M.K.,  
 Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 Thi-Ngoc H.P., Redder P., Schenk M.B., Theriault C., Tolstrup N.,  
 Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.,  
 "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840 (2001).  
 EMBL; AE006755; AA041639.1; -.  
 PIR; H90297; H90297.  
 InterPro; IPR003799; DUF196.  
 Pfam; PF02647; DUF196; 1.  
 TIGRFAMs; TIGR01573; cas2; 1.  
 Hypothetical protein; Complete proteome.  
 SEQUENCE 101 AA; 11936 MW; A7C338AD76202E17 CRC64;

Query Match 95.0%; Score 19; DB 17; Length 101;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5  
 ||||  
 20 VAEF 23

SULT 55  
 1R27 PRELIMINARY; PRT; 102 AA.

Q81R27  
 01-JUN-2003 (TReMBLrel. 24, Created)  
 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 Hypothetical protein.  
 B2232.  
 Bacillus anthracis (strain Ames).  
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 NCBI\_TaxID=198094;  
 [1]

SEQUENCE FROM N.A.  
 MEDLINE=22608414; PubMed=12721629;  
 Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,  
 Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
 Holtzapple E.K., Okestad O.A., Helgason E., Ristone J., Wu M.,  
 Kolonay J.F., Beaman R.J., Dodson R.J., Brinkac L.M., Gwinn M.,  
 DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
 Nelson W.C., Peterson J.D., Pop M., Khouiri H.M., Radune D.,

RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,  
 Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Niernan W.C.,  
 RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,  
 RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolato A.-B.,  
 RA Fraser C.M.;  
 "The genome sequence of Bacillus anthracis Ames and comparison to  
 RT closely related bacteria.";  
 RL Nature 423:81-86 (2003).  
 DR EMBL; AE017031; AAP26109.1; -.  
 DR TIGR; BA2232; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 102 AA; 11925 MW; 13B8E296C9A341B1 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 102;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 ||||  
 54 VAEF 57

Db

RESULT 56  
 Q722N2 PRELIMINARY; PRT; 103 AA.

ID Q722N2  
 AC Q722N2;  
 DT 01-OCT-2003 (TReMBLrel. 25, Created)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE BK150C2.6 (Putative novel protein similar to APOBEC1 (Apolipoprotein B  
 DE mRNA editing protein) and Phorbol (Fragment).  
 GN BK150C2.6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Matthews L.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL022318; CAB45276.1; -.  
 KW Lipoprotein.  
 FT NON\_TER 1 1  
 FT NON\_TER 103 103  
 SQ SEQUENCE 103 AA; 12146 MW; SDC969AE3ED348D9 CRC64;

Query Match 95.0%; Score 19; DB 4; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 ||||  
 46 VAEF 49

Db

RESULT 57  
 Q9F368 PRELIMINARY; PRT; 103 AA.

ID Q9F368  
 AC Q9F368;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Hypothetical protein SCO4420.  
 GN SCO4420 OR SC6F11.18.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,

```

A Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
A Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
A Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
A Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
A Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
A Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
A Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
A Hopwood D.A.;
T "Complete genome sequence of the model actinomycete Streptomyces
T coelicolor A3(2).";
L Nature 417:141-147(2002).
R EMBL: AL39120; CAC08429.1; -.
W Hypothetical protein; Complete proteome.
Q SEQUENCE 103 AA; 11387 MW; 55C2B80589B75B5 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 103;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
b 54 VAEF 57

RESULT 58
89MS9 PRELIMINARY; PRT; 103 AA.
C Q99MS9;
T 01-JUN-2003 (TrEMBLrel. 24, Created)
T 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
E B1r4113 protein.
N BLR4113.
S Bradyrhizobium japonicum.
C Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
C Bradyrhizobiaceae; Bradyrhizobium.
X NCBI_TaxID=375;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=USDA 110;
X MEDLINE=22484998; PubMed=12597275;
A Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
A Sasamoto S., Watanabe A., Idesawa K., Iriiguchi M., Kawashima K.,
A Kohara M., Matsumoto M., Shimo S., Teurtocka H., Wada T., Yamada M.,
A Tabata S.;
T "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
T Bradyrhizobium japonicum USDA110.";
L DNA Res. 9:189-197(2002).
R EMBL: AP005950; BAC49378.1; -.
W Complete proteome.
Q SEQUENCE 103 AA; 11104 MW; BF27CB9F90FBB723 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 103;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5
b 96 VAEF 99

RESULT 59
85EB5 PRELIMINARY; PRT; 104 AA.
C Q85EB5;
T 01-JUN-2003 (TrEMBLrel. 24, Created)
T 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
E Gp57.
S Mycobacteriophage Baryard.
C Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
X NCBI_TaxID=205860;
N [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=22592660; PubMed=12705866;
RA Pedulla M.L., Ford M.E., Houtz J.M., Karthikeyan T., Wadsworth C.,
RA Lewis J.A., Jacobs-Sera D., Falbo J., Gross J., Pannunzio N.R.,
RA Brucker W., Kumar V., Kandasamy J., Keenan L., Bardarov S.,
RA Kriakov J., Lawrence J.G., Jacobs W.R. Jr., Hendrix R.W.,
RA Hatfull G.F.;
RT "Origins of highly mosaic mycobacteriophage genomes.";
RL Cell 113:171-182(2003).
DR EMBL: AY129339; RAN02111.1; -.
SQ SEQUENCE 104 AA; 11074 MW; BB806EAC401B9FD0 CRC64;

Query Match 95.0%; Score 19; DB 9; Length 104;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
Db 59 VAEF 62

RESULT 60
Q8SVK0 PRELIMINARY; PRT; 105 AA.
ID Q8SVK0;
AC Q8SVK0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein ECU05_0770.
GN ECU05_0770.
OS Encephalitozoon cuniculi.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RX MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarar F.,
RA Prensier G., Barbe V., Peyretallade E., Brottier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi.";
RL Nature 414:450-453(2001).
DR EMBL: AL590445; CAD26596.1; -.
RW Hypothetical protein.
SQ SEQUENCE 105 AA; 11897 MW; 08059108C05D3BB9 CRC64;

Query Match 95.0%; Score 19; DB 5; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5
Db 77 VAEF 80

RESULT 61
Q856U1 PRELIMINARY; PRT; 105 AA.
ID Q856U1;
AC Q856U1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Gp3.
S Mycobacteriophage Corndog.
C Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=205975;

```

```

[1]
SEQUENCE FROM N.A.
MEDLINE=22592660; PubMed=12705866;
Peculia M.L., Ford M.E., Houtz J.M., Karthikeyan T., Wadsworth C.,
Lewis J.A., Jacobs-Sera D., Falbo J., Gross J., Panunzio N.R.,
Brucker W., Kumar V., Kandasamy J., Keenan L., Bardarov S.,
Kriakov J., Lawrence J.G., Jacobs W.R. Jr., Hendrix R.W.,
Hatfull G.F.;
"Origins of highly mosaic mycobacteriophage genomes.";
Cell 113:171-182(2003).
EMBL; AY129335; AA01935.1; -.
SEQUENCE 105 AA; 11955 MW; 7D0EC09F5F2AF2DD CRC64;

Query Match          95.0%; Score 19; DB 9; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5
|||||
41 VAEF 44

SULT 62
6231
Q96Z31 PRELIMINARY; PRT; 105 AA.
Q96Z31; 2001 (TrEMBLrel. 19, Created)
01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
Hypothetical protein ST2000.
ST2000.
Sulfolobus tokodaii.
Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
Sulfolobus.
NCBI_TaxID=111955;
[1]
SEQUENCE FROM N.A.
STRAIN=JCM 10545 / 7;
MEDLINE=21456156; PubMed=11572479;
Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
Ngai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
Oshima T., Kikuchi H.;
"Complete genome sequence of an aerobic thermophilic
Crenarchaeon, Sulfolobus tokodaii strain7.";
DNA Res. 8:123-140(2001).
EMBL; AP000988; BAB67095.1; -.
InterPro; IPR006783; Transposase_29.
Pfam; PF04693; Transposase_29; 2.
Hypothetical protein; Complete proteome.
SEQUENCE 105 AA; 11997 MW; AFE6ABD4DF2DECOE CRC64;

Query Match          95.0%; Score 19; DB 17; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5
|||||
29 VAEF 32

SULT 63
1530
P71530 PRELIMINARY; PRT; 106 AA.
P71530;
01-FEB-1997 (TrEMBLrel. 02, Created)
01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
NIFX.
Methanococcus maripaludis.

Archaea; Euryarchaeota; Methanococci; Methanococcales;
Methanococcaceae; Methanococcus.
NCBI_TaxID=39152;
[1]
SEQUENCE FROM N.A.
STRAIN=LL;
MEDLINE=96011360; PubMed=7592322;
Blank C.E., Kessler P.S., Leigh J.A.;
"Genetics in methanogens: transposon insertion mutagenesis of a
Methanococcus maripaludis nifH gene.";
J. Bacteriol. 177:5773-5777(1995).
[2]
SEQUENCE FROM N.A.
STRAIN=LL;
MEDLINE=97144542; PubMed=8990309;
Kessler P.S., McLarnan J., Leigh J.A.;
"Nitrogenase phylogeny and the molybdenum dependence of nitrogen
fixation in Methanococcus maripaludis.";
J. Bacteriol. 179:541-543(1997).
EMBL; U75887; AAC45519.1; -.
PIR; T10097; T10097.
DR InterPro; IPR003731; DUF153.
DR Pfam; PF02579; Nitro_Femo-Co_1.
SQ SEQUENCE 106 AA; 11862 MW; 4734D89CFED33F3B CRC64;

Query Match          95.0%; Score 19; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5
|||||
35 VAEF 38

RESULT 64
Q9BGL5 PRELIMINARY; PRT; 106 AA.
AC Q9BGL5; 2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Matrilysin (Fragment).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
[1]
SEQUENCE FROM N.A.
TISSUE=Endometrium;
RA Smith G.W., Ricke W.A., Cassar C.A., Smith M.F.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF267158; AAG59846.1; -.
DR HSP; P09237; 1MMQ.
DR MEROPS; M10.008; -.
DR GO; GO:0005578; C:extracellular matrix; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006026; Peptidase_M.
DR InterPro; IPR001818; Pept_M10A_M12B.
DR Pfam; PF00413; Peptidase_M10; 1.
DR PRINTS; PR00138; MATRINX.
DR SMART; SM00235; ZnMc; 1.
DR NON_TER 1.
FT NON_TER 106
SQ SEQUENCE 106 AA; 11578 MW; EBD271054928018 CRC64;

Query Match          95.0%; Score 19; DB 6; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5
|||||

```

b 6 VAEF 9

## RESULT 65

D Q23682 PRELIMINARY; PRT; 106 AA.

C Q23682;

T 01-JAN-1998 (T-EMBLrel. 05, Created)

T 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)

T 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)

E Cathepsin B-like cysteine proteinase (Fragment).

N 77123.13.

S Arabidopsis thaliana (Mouse-ear cress).

C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

C eucosids II; Brassicales; Brassicaceae; Arabidopsis.

X NCBI\_TaxID=3702;

N [1]

P SEQUENCE FROM N.A.

C STRAIN=cv. Columbia;

A Federspiel N.A., Conway A.B., Conway A.R., Davis K., Brendel V.,

A Palm C.J., Au M., Araujo R., Chung E., Kurtz D.B., Buehler E.,

A Dewar K., Feng J., Kim C., Li Y., Shinn P., Sun H., Oji O.,

A Osborne B., Shen Y.K., Toriumi M., Vyotskaia V., Yu G., Theologis A.,

A Ecker J., Davis R.W.;

T "Genomic sequence of Arabidopsis BAC T7123.";

L Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

R EMBL; U99959; AAC24377.1; -.

T NON TER 106 106

Q SEQUENCE 106 AA; 11743 MW; C5D5EA5897AD17AF CRC64;

## Query Match

Best Local Similarity 95.0%; Score 19; DB 10; Length 106;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5

|||||

75 VAEF 78

## RESULT 66

D Q9C1L1 PRELIMINARY; PRT; 107 AA.

C Q9C1L1;

T 01-JUN-2001 (T-EMBLrel. 17, Created)

T 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)

T 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)

E Putative thioredoxin G6G8.7.

S Neurospora crassa.

C Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

C Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

X NCBI\_TaxID=5141;

N [1]

P SEQUENCE FROM N.A.

C STRAIN=74-OR23-1A;

X MEDLINE=21135672; PubMed=11238395;

A Bean L.E., Dvorachek W.H. Jr., Braun E.L., Errett A., Saenz G.S.,

A Giles M.D., Werner-Washburne M., Nelson M.A., Natvig D.O.;

T "Analysis of the pdx-1 (enz-1/sno-1) region of the Neurospora crassa

genome. Correlation of pyridoxine-requiring phenotypes with mutations

in two structural genes.";

T Genetics 157:1067-1075 (2001).

L EMBL; AF309689; AAK07845.1; -.

R HSP; P80028; ITOF.

R GO; GO:0005489; F:electron transporter activity; IEA.

R GO; GO:0006118; P:electron transport; IEA.

R InterPro; IPR006662; Thiorid.

R InterPro; IPR006663; Thioridox\_dom2.

R Pfam; PF00085; Thiorid; I.

R PRINTS; PR00421; THIOREDOXIN.

R Redox-active center.

W SEQUENCE 107 AA; 11676 MW; CF4B6EAF85BE3776 CRC64;

Query Match 95.0%; Score 19; DB 3; Length 107;

Best Local Similarity 100.0%; Pred. No. 1.6e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5

|||||

9 VAEF 12

## RESULT 67

Q8IHV3

ID Q8IHV3 PRELIMINARY; PRT; 107 AA.

AC Q8IHV3;

DT 01-MAR-2003 (T-EMBLrel. 23, Created)

DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE Hypothetical protein.

GN PF11\_0423.

OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=36329;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=3D7;

RX MEDLINE=22255705; PubMed=12368864;

RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,

RA Carlton J.N., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,

RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,

RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,

RA Perlea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,

RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,

RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,

RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

RA Fraser C.M., Barrell B.;

RT "Genome sequence of the human malaria parasite Plasmodium

falciparum.";

RL Nature 419:498-511 (2002).

DR EMBL; AF014842; AAN36006.1; -.

KW Hypothetical protein.

SQ SEQUENCE 107 AA; 12714 MW; CA351258FCF46EFS CRC64;

## Query Match

Best Local Similarity 95.0%; Score 19; DB 5; Length 107;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5

|||||

40 VAEF 43

## RESULT 68

Q8NRX6

ID Q8NRX6 PRELIMINARY; PRT; 107 AA.

AC Q8NRX6;

DT 01-OCT-2002 (T-EMBLrel. 22, Created)

DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)

DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)

DE Uncharacterized ACR.

GN CGL0915.

OS Corynebacterium glutamicum (Brevibacterium flavum).

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.

OX NCBI\_TaxID=1718;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;

RA Nakagawa S.;

RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF005276; BAB98308.1; -.

DR InterPro; IPR007138; ABW.

DR Pfam; PF03992; ASW; 1.

DR Complete proteome.

KW

SEQUENCE 107 AA; 12373 MW; B5ACF23621078018 CRC64;

Query Match 95.0%; Score 19; DB 16; Length 107;

Best Local Similarity 100.0%; Pred. No. 1.6e+03; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5

|||||

22 VAEF 25

SULT 69

P082

PRELIMINARY; PRT; 109 AA.

Q9PU82

01-MAY-2000 (TREMBlrel. 13, Created)

01-MAY-2000 (TREMBlrel. 13, Last sequence update)

01-JUN-2003 (TREMBlrel. 24, Last annotation update)

Retinol binding protein (Fragment).

Crocodylus niloticus (Nile crocodile) (African crocodile).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Crocodylidae; Crocodylinae; Crocodylus.

NCBI\_TaxID=8501;

[1]

SEQUENCE FROM N.A.

TISSUE=Liver;

MEDLINE=20022983; PubMed=10555283;

Hughes S., Zelus D., Mouchiroud D.;

"Warm-blooded isochore structure in Nile crocodile and turtle.";

Mol. Biol. Evol. 16:1521-1527(1999).

EMBL; AJ011392; CAB56418.1; -.

HSP; P02753; IRBP.

GO; GO:0005215; F:transporter activity; IEA.

GO; GO:0006810; P:transport; IEA.

InterPro; IPR002345; Lipocalin.

InterPro; IPR000566; Lipocln\_cyfPABP.

Pfam; PF00061; lipocalin; 1.

PRINTS; PR00179; LIPOCALIN.

NON\_TER 1

NON\_TER 109

SEQUENCE 109 AA; 12552 MW; ODCAE9389591A23 CRC64;

Query Match

Best Local Similarity 95.0%; Score 19; DB 13; Length 109;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5

|||||

25 VAEF 28

SULT 70

3333

PRELIMINARY; PRT; 109 AA.

O53333

01-JUN-1998 (TREMBlrel. 06, Created)

01-JUN-1998 (TREMBlrel. 06, Last sequence update)

01-OCT-2003 (TREMBlrel. 25, Last annotation update)

Hypothetical protein RV3183.

RV3183 OR M7V014.27 OR M73275.

Mycobacterium tuberculosis.

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium.

NCBI\_TaxID=1773;

[1]

SEQUENCE FROM N.A.

STRAIN=H37RV;

MEDLINE=98295987; PubMed=9634230;

Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,

Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

"Deciphering the biology of Mycobacterium tuberculosis from the

complete genome sequence.";

Nature 393:537-544 (1998).

[2]

SEQUENCE FROM N.A.

STRAIN=CDC 1551 / Oshkosh;

Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

Koionay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

Decher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

Bishai W.;

"Whole genome comparison of Mycobacterium tuberculosis clinical and

laboratory strains.";

Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

EMBL; AL021646; CAA16648.1; -.

EMBL; AE007140; AAK47615.1; -.

PIR; E70949; E70949.

TIGR; MT3275; -.

TuberculList; RV3183; -.

GO; GO:0003677; F:DNA binding; IEA.

InterPro; IPR001387; HTH 3.

Pfam; PF01381; HTH 3; 1.

SMART; SM00530; HTH\_XRE; 1.

Hypothetical protein; Complete proteome.

SEQUENCE 109 AA; 11802 MW; A66DBE160D79613A CRC64;

Query Match

Best Local Similarity 95.0%; Score 19; DB 16; Length 109;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5

|||||

97 VAEF 100

RESULT 71

Q7TX28

PRELIMINARY; PRT; 109 AA.

AC Q7TX28;

01-OCT-2003 (TREMBlrel. 25, Created)

01-OCT-2003 (TREMBlrel. 25, Last sequence update)

01-OCT-2003 (TREMBlrel. 25, Last annotation update)

Possible transcriptional regulatory protein.

GN M3209.

Mycobacterium bovis.

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium.

NCBI\_TaxID=1765;

[1]

SEQUENCE FROM N.A.

STRAIN=AF2122/97.

MEDLINE=22709107; PubMed=12788972;

Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,

Pryor M., Duthoy S., Grondin S., Lacroix C., Monsenge C., Simon S.,

Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,

Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;

"The complete genome sequence of Mycobacterium bovis.";

Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).

EMBL; BX248345; CAD95301.1; -.

Complete proteome.

SEQUENCE 109 AA; 11802 MW; A66DBE160D79613A CRC64;

Query Match

Best Local Similarity 95.0%; Score 19; DB 16; Length 109;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VAEF 5

|||||

97 VAEF 100

RESULT 72  
 11343 PRELIMINARY; PRT; 110 AA.  
 C O11343;  
 D 01-JUL-1997 (TrEMBLrel. 04, Created)  
 T 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 T 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 E X1-5 protein (Fragment).  
 N X1-5  
 S Molluscum contagiosum virus subtype 1 (MCV1)  
 C Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 C Molluscipoxvirus.  
 X NCBI\_TaxID=10280;  
 [1]  
 P SEQUENCE FROM N.A.  
 A Moratilla M., Agronayor M., Nunez A., Funes J.M., Varas A.J.,  
 A Lopez-Esteban J.L., Esteban M., Martin-Gallardo A.,  
 T "A Random DNA Sequencing, Computer-Based Approach for the Generation  
 T of a Gene Map of Molluscum Contagiosum Virus."  
 L Virus Genes 0:0-0(1997).  
 R EMBL; U86919; AAB57977.1; -.  
 T NON\_TER 1 1  
 T NON\_TER 110 110  
 T SEQUENCE 110 AA; 11856 MW; 6AB9CE25FC3DB780 CRC64;

Query Match 95.0%; Score 19; DB 12; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 VAEF 5  
 b 9 VAEF 12

RESULT 73  
 88EG3 PRELIMINARY; PRT; 110 AA.  
 D Q88EG3  
 C Q88EG3;  
 T 01-JUN-2003 (TrEMBLrel. 24, Created)  
 T 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 E YCII-related domain protein.  
 N P4502  
 S Pseudomonas putida (strain KT2440).  
 C Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 C Pseudomonadaceae; Pseudomonas.  
 X NCBI\_TaxID=160488;  
 [1]  
 P SEQUENCE FROM N.A.  
 A MEDLINE=22423060; PubMed=12534463;  
 A Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,  
 A Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,  
 A Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,  
 A Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,  
 A Chris Lee P., Holtzapfe E., Scanlan D., Tran K., Moazzes A.,  
 A Uterback T., Rizzo M., Lee K., Kosack D., Woestl D., Wedler H.,  
 A Lauber J., Stjepandic D., Hobeisel J., Straetz M., Heim S.,  
 A Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,  
 A Fraser C.M.;  
 T "Complete genome sequence and comparative analysis of the  
 T metabolically versatile Pseudomonas putida KT2440.";  
 L Environ. Microbiol. 4:799-808(2002).  
 R EMBL; AE016790; AAN70076.1; -.  
 R TIGR; PP4502;  
 R InterPro; IPR005545; YCII.  
 R Pfam; PF03795; YCII; 1.  
 W Complete proteome.  
 Q SEQUENCE 110 AA; 11775 MW; 774B31C0D925A01E CRC64;

Query Match 95.0%; Score 19; DB 16; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 Db 73 VAEF 76  
 RESULT 74  
 Q8U296 PRELIMINARY; PRT; 111 AA.  
 ID Q8U296  
 AC Q8U296;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein PF0943.  
 GN PF0943.  
 OS Pyrococcus furiosus.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=2261;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Yc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
 RT "The complete sequence of the Pyrococcus furiosus genome."  
 RT Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE010208; BA181067.1; -.  
 DR InterPro; IPR007842; HEPN.  
 DR Pfam; PF05168; DUF712; 1.  
 DR PROSITE; PS0910; HEPN; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 111 AA; 13379 MW; C6A005A6B2B6CA4E CRC64;

Query Match 95.0%; Score 19; DB 17; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 Db 48 VAEF 51  
 RESULT 75  
 Q86PL2 PRELIMINARY; PRT; 112 AA.  
 ID Q86PL2  
 AC Q86PL2;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Nuclear receptor NHR-1 (Fragment).  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 [1]  
 P SEQUENCE FROM N.A.  
 A Robinson-Rechavi M., Maina C.V., Gissendanner C.R., Laudet V.,  
 A Studer A.;  
 RT "Explosive lineage-specific expansion of the orphan nuclear receptor  
 RT HNF4 in nematodes.";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AV204162; AAC39166.1; -.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 DR InterPro; IPR008946; Str\_ncl\_receptor.  
 KW Receptor.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 112 AA; 12826 MW; 44FC60A4E3744791 CRC64;

Query Match 95.0%; Score 19; DB 5; Length 112;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VAEF 5  
 Db 73 VAEF 76

13 VAEF 16

arch completed: May 24, 2004, 17:41:06  
b time : 37.7143 secs